Identifying Kinship Cues from Facial Images

POLITECNICO DI TORINO

SCUOLA DI DOTTORATO

Dottorato in Automatica e Informatica – XV ciclo

Tesi di Dottorato

Identifying Kinship Cues from Facial Images

Tiago Figueiredo Vieira
matricola: 172569

Coordinatore del corso di dottorato
prof. Pietro Laface

Novembre 2013

POLITECNICO DI TORINO
DEPARTMENT OF CONTROL AND COMPUTER ENGINEERING

IDENTIFYING KINSHIP CUES FROM FACIAL IMAGES

by

TIAGO FIGUEIREDO VIEIRA

Thesis submitted to Politecnico di Torino as a partial fulfillment of the requirements for the degree of Philosophy Doctor in Control and Computer Engineering.

ADVISOR: ANDREA GIUSEPPE BOTTINO, Ph. D.
CO-ADVISOR: EDUARDO FONTANA, Ph. D.

Torino, november 2013.

Folha de aprovação

To my beloved son, Lucas.

I would like to thank:

My parents Amilton and Carmem for all their love and support.

My wife Marcelle for her love, patience and partnership.

My son Lucas, for teaching me the meaning of unconditional love.

Abstract of Thesis presented to POLITO as a partial fulfillment of the requirements for the degree of Doctor in Control and Computer Engineering.

Identifying Kinship Cues from Facial Images

Tiago Figueiredo Vieira

November/ 2013

Supervisor(s): Andrea Giuseppe Bottino, Ph. D., Eduardo Fontana, Ph. D.

Area of Concentration: Computer Engineering.

Keywords: Kinship Verification, Support Vector Machines, Feature Selection.

Number of Pages: 88.

The investigation of human face images is ubiquitous in pattern analysis/ image processing research. Traditional approaches are related to face identification and verification but, several other areas are emerging, like age/ expression estimation, analysis of facial similarity and attractiveness and automatic kinship recognition. Despite the fact that the latter could have applications in fields such as image retrieval and annotation, little work in this area has been presented so far. This thesis presents an algorithm able to discriminate between siblings and unrelated individuals, based on their face images. In this context, a great challenge was to deal with the lack of a benchmark in kinship analysis, and for this reason, a high-quality dataset of images of siblings' pairs was collected. This is a relevant contribution to the research community and is particularly useful to avoid potential problems due to low quality pictures and uncontrolled imaging conditions of heterogeneous datasets used in previous researches. The database includes frontal, profile, expressionless and smiling faces of siblings pairs. Based on these images, various classifiers were constructed using feature-based and holistic techniques to investigate which data are more effective for discriminating siblings from non-siblings. The features were first tested individually and then the most significant face data were supplied to a unique algorithm. The siblings classifier has been found to outperform human raters on all datasets. Also, the good discrimination capabilities of the algorithm is tested by applying the classifiers to a low quality database of images collected from the Internet in a cross-database experiment. The knowledge acquired from the analysis of siblings fostered a similar algorithm able to discriminating parent-child pairs from unrelated individuals. The results obtained in this thesis have impact in image retrieval and annotation, forensics, genealogical research and finding missing family members.

Contents

List of Figures viii

List of Tables xii

1 Introduction 1

2 Literature Review 5

I Classification of Siblings 10

3 Image Databases 11

3.1 High Quality database 11

3.2 Low Quality database 14

3.3 Database normalization 15

3.3.1 Detection of facial landmarks 15

3.3.2 Geometric normalization 16

3.3.3 Background removal 17

3.3.4 Intensity normalization and face cropping 17

4 Extraction of Features 19

4.1 Geometric features 23

4.2 Holistic attribute (PCA) 25

4.3 Texture descriptors 26

4.3.1 Rotation-Invariant Co-occurrence of Local Binary Patterns (RIC-LBP) 26

4.3.2 Rotation-Invariant Gabor Feature (RIGF) 30

4.3.3 Color descriptors 30
4.4 Features summary 32

5 Classification and Feature Selection 33
5.1 Support Vector Machines (SVM) 33
5.1.1 Intuition 33
5.1.2 Functional and geometric margins 34
5.1.3 The optimal margin classifier 37
5.1.4 Lagrange duality 38
5.1.5 Optimal margin classifiers 41
5.2 Feature Selection 43

6 Results and Discussion 46
6.1 Preliminary observations 46
6.1.1 Analysis of facial distances 47
6.1.2 Siblings classification using only *eigenfaces* 48
6.1.3 Scale-Invariant Feature Transform (SIFT) 49
6.1.4 Color descriptors 50
6.2 Discriminating between siblings and non-siblings 52
6.3 Human classification 53
6.4 Commercial software 55
6.5 Automatic siblings classifier 56

II Classification of Parent-Child Pairs 67

7 Parent-Child Classification 68
7.1 Parent-child database 69
7.2 Weber Local Descriptor (WLD) 71
7.3 Results of parent-child classification 75

8 Siblings Classification Revisited 78
8.1 New set of features 78
8.2 Classification results 79
8.3 Selection of best features 80

8.4 Correct and incorrect classifications 81
8.4.1 Processing times 84

9 Conclusion and Future Work **86**
9.1 Future work 87

Acronyms **89**

Bibliography **91**

List of Figures

2.1 Illustration on how to compute the kinship coefficient (relatedness) given by $r = (1/2)^n$, where n is the smallest number of lines connecting two individuals. 6

2.2 Subject asked to choose between two political nominees. 6

2.3 Relation between kinship assessment performed by humans and different degrees of relatedness: Id - same individual; Sb - siblings; Gp - grandparent-child pairs; Au - aunt/uncle, nephew/niece pairs; Cs - cousins. 7

2.4 Examples of positive and negative samples of parent-child pairs from the work of Fang *et al.*. 8

3.1 Examples of images of family groups belonging to HQ_{faces}: (a) three siblings, but the girl's image is discarded due to hair occlusion; and, (b) faces of three pairs of siblings in neutral frontal, smiling frontal and smiling profile poses. 12

3.2 Distribution of ages of participants. 13

3.3 Four pairs of siblings belonging to the LQ_{faces} image database. 15

3.4 Landmarks detected on different image types. The numbers in the pictures pinpoint the elements chosen for geometric normalization. 16

3.5 Examples of normalized and cropped grayscale images of an individual in *HQ-fps*. 18

4.1 Feature extraction illustration. 21

4.2 Reference segments, obtained from Delaunay Triangulation, for frontal (left) and profile (right) images. 24

4.3 Examples of *eigenfaces* for set *HQ-fps*. 26

4.4 Illustration of how to compute a Local Binary Pattern (LBP). Resulting pattern is 00100010_2, or 34_{10}. 27
4.5 Illustration of how to compute CoALBP features from an image. 28
4.6 Illustration on how to compute the RIC-LBP feature. 29
4.7 Example of RIGF feature computed onto an original and rotate texture. 30

5.1 Intuition on how Support Vector Machines work: (a) randomly generated data samples of two different classes on 2D space; (b) samples mapped into a higher dimensional space (3D), and (c) linear separation of the data using a plane. 34
5.2 Separation of two classes by a hyperplane defined by parameters $\mathbf{w}, b$. . . 35
5.3 Sequential Forward Selection strategy: an example (for C-SIFT attribute and *HQ-fps* dataset) of the behavior of the classification accuracy on different iterations. 43
5.4 Illustration of the Feature Selection (FS) procedure. 44
5.5 Outline of the sibling recognition process. 45

6.1 Optimization of SVM parameters yielding a 63.2% accuracy in siblings classification considering only the euclidean distance between nose tip and mouth. 48
6.2 Most relevant *eigenfaces* for the siblings classification problem in decreasing order of importance from left to right and top to bottom: (a) best *eigenfaces* for *HQ-f* PDS, and; (b) best *eigenfaces* for LQ_{faces} dataset. . . 49
6.3 SVM classification for datasets composed of sift and surf descriptors computed using different window sizes. (a) Frontal expressionless image of one individual from *HQ-f*. (b) SVM accuracies for SIFT and SURF using different window sizes. (c) Expressionless profile image from *HQ-fp*. (d) SVM accuracies for SIFT and SURF computed with different window sizes. 51
6.4 Features selected after the two-step feature selection applied to the SIFT obtained from frontal HQ images. 51
6.5 Features selected after the two-step feature selection applied to the Local Color Descriptor obtained from frontal HQ images. 52

6.6 Example images presented in the questionaire answered by members of a human panel; (a) positive pair belonging to set *HQ-fps*; (b) negative pair belonging to set *HQ-fp*; and (c) question to be answered with YES/ NO options. 54
6.7 Histograms of scores provided by FaceVACs-SDK for pairs of siblings and non-siblings in: (a) HQ_{faces} and (b) LQ_{faces} (on the bottom). 55
6.8 ROC curves obtained from applying FaceVACS to both LQ_{faces} and HQ_{faces}. 56
6.9 Illustration of the accuracy values of individual attributes from Table 6.2. 59
6.10 Illustration of the accuracy values of grouped attributes from Table 6.2. . 60
6.11 Proportions of chosen features selected from Sequential Forward Selection (SFS) applied to ALL attribute group. 62
6.12 Results of other classification techniques. 63

7.1 Database of parent-child pairs collected over the internet by Fang *et al.*. . 69
7.2 Examples of three parent-child pairs belonging to PC_{faces}. Images on the left show originals with annotated landmarks, whereas images on the right show the normalized faces. 71
7.3 Two examples of the computation of the Weber Local Descriptor components: (a) and (d): original images; (b) and (e): differential excitations; (c) and (f): gradient orientations. Images (b), (c), (e) and (f) are scaled to $[0, 255]$ for visualization purpose. 73
7.4 Weber Local Descriptor matrix. 74
7.5 The final Weber descriptor vector $\mathbf{h} \in \Re^{2880}$ (or histogram) computed from the image shown in Figure 7.3a. 74
7.6 Classification accuracies of individual and combined features for parent-child pairs. 76
7.7 Feature selection applied to (a) SVM and (b) RDF: distribution of feature variables per type for different attribute groups. 77

8.1 Classification accuracies for individual and grouped attributes for PDSs *HQ-f*, *HQ-fp*, and *HQ-fps* at left, center and right, respectively. 81
8.2 Proportions of features selected from ALL attribute group computed from *HQ-fps*. 82

8.3 Comparison between previous and current accuracies in siblings classification. 82

8.4 Examples of correctly and erroneously classified samples. 83

List of Tables

3.1 Summary of Individual Datasets (IDS). 14

3.2 Averages and standard deviations of age differences between siblings who participated in the photo sessions. 14

3.3 Reference positions used for geometric normalization. The image coordinate system has its origin on the top left and the x and y axes are, respectively, horizontal and vertical. 17

3.4 Characteristics of the reduced standard areas. 17

4.1 Modeling of variations in light intensity and color. 31

4.2 Attributes dimensions. 32

6.1 Classification results from the human panels experiment. 54

6.2 Classification results, showing for each PDS and for each attribute or attribute group, the classification accuracy (Acc) and the percentage of features selected by the FS algorithm (NFS). Higher accuracies are highlighted with brighter backgrounds. 58

6.3 Classification accuracies (using ALL attributes group) applying SVM to (i) the initial feature set (NOFS), (ii) the reduced feature set extracted by mRMR (FS1) and (iii) the final feature set obtained by combining mRMR and SFS (FS2). 60

6.4 Correlation between the classification variable and age differences. 62

6.5 Classification accuracies per pose of faces belonging to set *HQ-fps*. 63

6.6 Comparison of automatic and human classification of different datasets. . 64

6.7 Cross and within database classification accuracies on LQ_{faces} with ALL attribute group and different color descriptors. 65

7.1 Accuracy results. For each attribute and each classification algorithm, we show the percentage of correct classifications and, in brackets, the optimal number of variables selected by the FS process. 75

8.1 Attributes dimensions. 79

8.2 Classification results, showing for each PDS and for each attribute or attribute group, the classification accuracy (Acc) and the percentage of features selected by the FS algorithm (NFS). 80

Chapter 1

Introduction

Human faces convey much information to other human beings and have been much investigated in the development of pattern analysis/ machine and learning techniques. For instance, there are the identification and verification approaches, whose implementations are ubiquitous in everyday life and have well established applications and algorithms. The former focuses on identifying an individual among a database of many people, whilst the latter aims at verifying whether the individual is who he says he is. In both cases, the general idea is to perform the matching between two similar faces. Among many others, one typical application is security.

In face image analysis, several other areas are emerging, such as: (i) affective computing [1], which consists in improving the human-computer interface by automatically recognizing the users' emotional state; (ii) age estimation [2], with applications in forensics, and; (iii) analyzing attractiveness [3], for surgical/ orthodontics planning. Each with its own idiosyncrasies, possible applications and difficulties.

The problem of identifying facial kinship clues with objective pattern analysis and image processing techniques has recently attracted the interest of researchers. Possible applications are historic and genealogic research, automatic management and labeling of image databases, forensic science, finding missing family members, etc. Many applications might be conceived if the machine becomes able to discriminate kins from unrelated people based on the inspection of their photographs. However, very few works have been presented so far and, in particular, the field lacks a benchmark, to foster further investigations in this area.

Automatic kinship recognition is an inherently challenging and, for this very reason,

a much interesting topic to be investigated. When analyzing the facial images of two individuals, the computer must tell whether those subjects are related or not. In addition, different degrees of kinship might be detected, as granparent-child, parent-child, siblings and so forth. In this case, several complications emerge due to variations in gender, age and smaller similarity between distant kins. This characterizes the field of automatic kinship recognition as a much broader and difficult problem than traditional face identification.

Taking this into consideration, this work presents an original investigation in the detection of kinship cues from human faces. In the direction of gradually increasing difficulty, only siblings were considered *a priori* and are treated in Part I. In this case, difficulty is expected mostly due to difference in gender, since the age difference between siblings are lower than more distant kins. An additional difficulty may arise from the possible presence of twins, which might confuse the classification with identification approach.

On a second stage, parent-child pairs were considered but were not mixed with the siblings classification. In this case, more difficulty emerge due to higher age differences, in addition to the gender problem. The parent-child classification is treated in Part II, where the knowledge gained in the previous investigation of siblings classification is applied, conceiving a similar, but slightly different approach. This strategy of first studying siblings prior to parent-child pairs was used in order to gradually obtain insights into the new problem of kinship recognition, given the lack of previous works to serve as comparison. Moreover, although the algorithms used to discriminate siblings and parent-child pairs are fairly similar, they are not strictly the same. Therefore, the parent-child classification problem is presented in a separate part (*cf.* Part II).

The general idea used in both cases consists in; (i) extracting several facial attributes from the normalized face images of two individuals; (ii) representing these facial traits mathematically; and finally (iii) trying to discriminate between the representative vectors of two kins and two unrelated individuals. As described before, each of these steps is performed slightly differently for siblings and for parent-child pairs.

Firstly, to assert that the investigation of siblings was not affected by image acquisition artifacts and database heterogeneity, a high quality database of siblings was assembled. Such images were collected through professional photo sessions performed in the Politecnico di Torino, where students and employees attended with their family members. The

images possess high quality and uniformity, and people were photographed in frontal and profile poses, with neutral and smiling expressions. This database is a particularly important contribution to the research community, specially in this new area of kinship recognition. All databases used in this work are available for academic purposes to other researches [4].

Secondly, several attributes were extracted from all individual faces and represented by mathematical variables. These attributes can be; (i) geometric; (ii) holistic; and (iii) textural and, although most of them were used in face analysis, some of them have not been reported in the literature as effective for facial recognition and, therefore, they are a contribution to the research community. In addition, prior to conceiving new attribute extraction techniques for kinship recognition, State of the Art (SoA) algorithms must be tested, in order to provide a reference basis for further investigation, since there are just a few results in the literature in this specific topic. Considering the lack of high-quality databases of parent-child image pairs, the investigation of parent-child cues was performed using low quality images collected from the internet by previous researches.

Last but not least, the feature vectors of two individuals obtained from their corresponding faces are compared, yielding the representative vector of a couple. The classification stage is then responsible for comparing the representative vectors of two couples and discriminating whether they correspond to kins or not. To this end, state of the art classification techniques were used with feature selection stages to improve the performance. In particular, results have shown that the combination of features of different nature provide higher performance than when the features are used individually. In addition, the higher the heterogeneity of the features used, the better the performance.

All image databases used in this work were also analyzed by human raters, who were asked to tell, based on visual inspection, whether two face photographs belonged to kins or not. Human panel experiments are commonly used in pattern analysis to provide a reasonable basis for comparison with the learning algorithms, given their heuristic nature. If the machine performs better than humans, it is an indication that it has statistical relevance. Simulations performed during the development of this thesis have shown that the machine is indeed capable of identifying siblings and parent-child pairs more accurately than humans, which is also a relevant contribution.

A preliminary version of this work was presented in [5], where we analyzed the use

of holistic techniques to identify siblings. In this work we present a deeper insight into the problem and a larger set of experiments supporting the findings, relative to those published in [6]. A very similar but slightly different methodology was applied to the discrimination between parent-child pairs and unrelated individuals, which was published in [7].

This thesis is organized as follows. A review of the current state of the art in automatic kinship recognition is provided in Chapter 2. Parts I and II present the methods and results to automatically discriminate Siblings and Parent-Child pairs. Chapter 3 describes how the images composing different siblings databases were acquired and organized. Chapter 4 details the proposed features to describe relatedness and their rationale. The proposed method for automatic sibling classification is given in Chapter 5. Experimental results of siblings analysis are presented and discussed in Chapter 6. Chapter 7 presents the classification of parent-child pairs, performed with knowledge learned from the experiments with siblings. Chapter 8 shows the problem of siblings classification revisited, using knowledge acquired from the parent-child classification problem. This structure of the thesis was adopted to maintain coherence with results published in the literature, as well as the chronology of the work developed. Finally, conclusions are drawn in Chapter 9.

Chapter 2

Literature Review

The recognition of kins has been studied in diverse fields such as biology, psychology and sociology, to name a few. In the field of evolutionary biology, Hamilton [8] put forward the theory of "inclusive fitness", which deals with the possible evolution of characters benefiting or not close relatives. According to this theory, from which derives a genetic model for the analysis of interactions between relatives of the same generation, one individual can behave either (*i*) selfishly, where it gains fitness from others; or (*ii*) altruistically, where it loses fitness to others. In addition, what determines the subjects' behavior is the amount of fitness quantity being exchanged and the perception of kinship between individuals. Hamilton hypothesizes that, on average, one individual is willing to perform an altruistic action to evolve a sibling at least twice the corresponding loss to the self. And similarly, "siblings deprive one another of reproductive prerequisites provided they can themselves make use of at least one half of what they take; individuals deprive half-siblings of four units of reproductive potential if they can get personal use of at least one of them; and so on." Also, it is worthwhile to deprive a large number of distant relatives in order to extract a small reproductive advantage. In simple words, Hamilton observed that recognizing kinship and also the degree of relatedness is very relevant to social behavior of animals and humans.

Simply put, the degree of relatedness can be obtained as illustrated in Figure 2.1, which represents the relations between parent-children and siblings by solid and dotted lines, respectively. The relatedness between two individuals is given by $r = (1/2)^n$, where n is the smallest number of lines connecting the two individuals, solid or not. For instance, the kinship coefficient between parent-children and siblings is 1/2, whilst

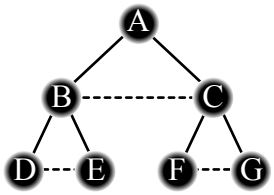


Figure 2.1: Illustration on how to compute the kinship coefficient (relatedness) given by $r = (1/2)^n$, where n is the smallest number of lines connecting two individuals.

between subjects D and G is 1/8.

Another example of the significance of kinship recognition was presented by Bailenson *et al.* [9]. They observed that the facial similarity between voters and candidates might influence the decision of the former. This resulted from different experiments where, in general, each participant (potential voter), unaware of the manipulation imposed on the images, was asked to choose between the photographs of two candidates, as shown in Figure 2.2. The first campaigner would have his mugshot morphed with another randomly chosen voter, whilst the second one would be morphed with the subject whom the question was being asked to. Experiments showed in general that, voters prefer candidates with whom they share facial similarities and this is especially true amid weak partisans and unfamiliar nominees.

Research highlights that human beings have a natural capability of recognizing kinship relationships between unknown individuals [10] and that the processes underlying kinship and identity recognition are probably different [11, 12]. Moreover, the performance in recognizing kins decreases for smaller degrees of relatedness as shown in Figure 2.3.

The facial characteristics shared by kins can be very different. For instance, although siblings, on average, have 50% of their genes in common, there are also extreme cases of siblings that do not seem to have any similarity between each other while conversely

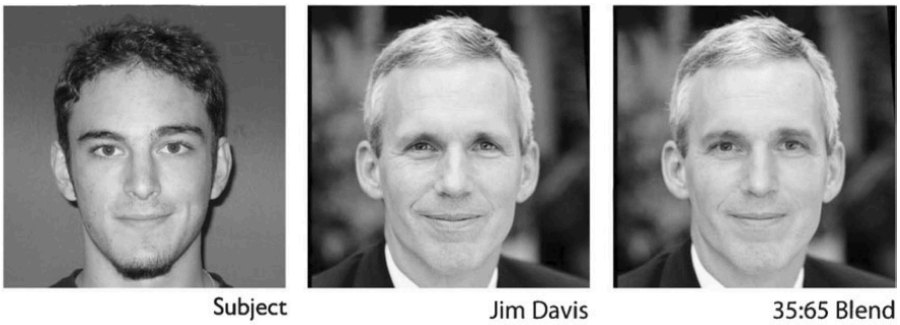


Figure 2.2: Subject asked to choose between two political nominees.

there are identical twins, whose individuality in some cases cannot be perceived by a simple inspection of their photographs [13]. Human science researchers have investigated the ability of human raters to recognize kinship from face images, attempting to identify the facial features providing kinship clues. Kaminski *et al.* [10] reported a 66% correct classification of kinship for siblings, using a data set of face images shot in uncontrolled conditions. By comparison, the same raters did not exceed 73% of kinship assessment when shown two images of the same person. Dal Martello and Maloney [14], on the basis of a high quality data set of children images, found that the upper part of the face carries more kinship clues.

The first computer analysis of facial features for a set of parent/child images was presented in 2010 by Fang *et al.* [15]. A database containing 150 semi-frontal image pairs, collected from the Internet, was analyzed, as illustrated in Figure 2.4. Twenty-two facial features and small windows surrounding feature points were extracted according to the Pictorial Structure Model. The k-Nearest Neighbors (KNN) and Support Vector Machines (SVM) classification schemes provided accuracies of 70.67% and 68.60%, respectively. The average classification accuracy of 67.19% for the same dataset was achieved by a panel of human raters.

In Somanath *et al.* [16] the problem of verifying kinship on a dataset of 43 child-parent and 26 sibling frontal image pairs, at a low resolution and shot in various lighting conditions, was addressed by training one classifier for each of these two classes using metric learning [17]. Feature vectors representing each pair were obtained combining

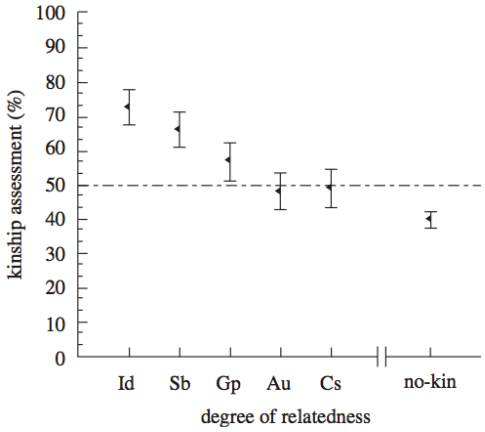


Figure 2.3: Relation between kinship assessment performed by humans and different degrees of relatedness: Id - same individual; Sb - siblings; Gp - grandparent-child pairs; Au - aunt/uncle, nephew/niece pairs; Cs - cousins.

local image descriptors, Gabor wavelets and intensity related information. Test samples were labeled from the classifier providing the highest confidence, achieving accuracies of 80% and 75%, respectively, for parent-child and siblings.

Starting from the observation that the similarity between facial images of parent and child is greater when the parent is younger, Xia *et al.* [18] proposed an extended Transfer Subspace Learning (TSL) approach to simplify the identification of parent-child pairs. TSL is meant to improve the knowledge of a new task with the transfer of knowledge learnt from a similar, but easier task. Their key idea is to introduce an intermediate class (containing images of parents in youth) that is close to both target classes (children and their parents) and to learn an optimal discriminative subspace between target classes through TSL. Classification accuracy reached 60% on a dataset collected over the Internet, approximately 3% higher than traditional TSL. An improved version of this method [19] achieved an accuracy of 69.7% on the same dataset, outperforming human classification (56%).

Finally, Guo and Wang [20] identified the facial familial traits shared by a pair of family members, as the building blocks of an automatic system for kinship verification. Given a labelled training set, familial traits are identified by comparing pairs of corresponding features (*e. g.* eyes, nose and mouth), and then used to compute the probability of a new feature pair being familial or not. For two individuals, the probabilities associated to their feature pairs are stochastically combined to make a decision. The overall accuracy, obtained on a dataset of almost frontal images in unconstrained illumination conditions, is 75%.

This work is aimed at performing a comprehensive exploration of the computer identi-

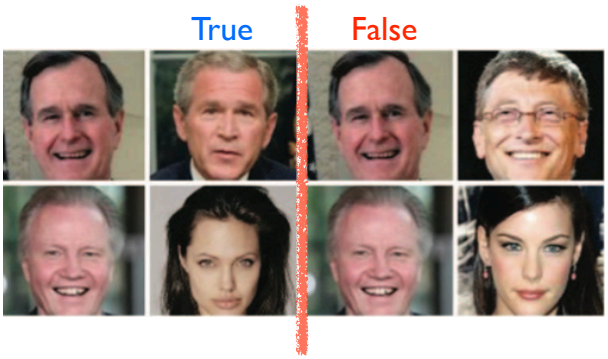


Figure 2.4: Examples of positive and negative samples of parent-child pairs from the work of Fang *et al.*.

fication of pairs of sibling' images with several pattern recognition techniques. The main contribution of our work is twofold. First, in order to avoid problems related to the low quality of the images used by other researchers, we constructed a dataset of high quality images specifically collected for this research. These images depict siblings in frontal and profile position, with and without expression and shot in controlled lighting conditions. Second, we analyzed the discriminative capabilities of different facial attributes to tell siblings from non-siblings. These attributes, related to holistic and feature-based classification techniques, were first used singularly in various classifiers, in order to investigate which data are more effective for our problem. Then, they were combined together into a more effective classifier. Image pair classification is based on the Support Vector Machines algorithm and on the integration of a two-step Feature Selection (FS) process, which led to improvements in the classification accuracy. Given the diffusion of face identification software, we also tested the ability of one commercial package to identify sibling pairs.

Several experiments were performed on our high quality database. The human capability to discriminate siblings was also verified on the same data. One important result is that the classifier combining all the face data consistently outperforms the recognition capabilities of human panels. The image pair classifiers were also tested with heterogeneous image sets, showing their generalization properties.

Part I

Classification of Siblings

Chapter 3

Image Databases

The first difficulty related to the investigation of kin recognition arises from the lack of good image databases to work on. Although not ideal, one possible solution is collecting images of public figures/ celebrities who are known to share degrees of kinship. From such approach many problems arise that might seriously affect the investigation, such as (*i*) non-uniform illumination; (*ii*) head pose not strictly frontal; (*iii*) expression not always neutral and often with smile; (*iv*) variety in background patterns/colors; and (*v*) usage of makeup by most celebrities, which affects color analysis. Not to mention difficulties inherent to analysis of kinship as different age range and ethnicity.

Taking this into consideration, a database of high quality images (resolution of 4256×2832 pixels) was collected from students and employees of the Politecnico di Torino, as explained in Section 3.1. After that, to analyze the generalization capabilities of the classification approach, a test database composed of low quality images of siblings' pairs (celebrities) from the Internet was assembled. Its properties are presented in Section 3.2. After the databases descriptions, Section 3.3 explains all pre-processing stages implemented on each database.

3.1 High Quality database

In order to avoid unwanted interference of artifacts due to the aforementioned causes, a database of high quality facial images of siblings was acquired. This was done by luring several students and employees of the Politecnico di Torino and respective siblings to attend professional photo sessions. People were asked not to wear makeup and to bring

their brothers and sisters, regardless of age and gender (see Figure 3.1a). The more the family members attending a photo session the higher the chance the person vinculated to the university responsible for bringing them had to win one of two possible prizes (a smartphone or a tablet PC). In this way, several photographs of different family groups were acquired under a strictly controlled environment, *i. e.*, normalized illumination and uniform, green background.

For an extended analysis of facial properties, some individuals were photographed under different poses and/ or expressions. The former could be frontal or profile whilst the latter could be either neutral or smiling. Hence, each subject is represented by one (neutral frontal), two (expressionless frontal and profile) or all four possible combinations, *i. e.*, neutral and smiling frontal and profile images.

During twenty days of photo sessions carried out in march 2011, 584 photographs were taken of 208 caucasian subjects with ages varying from 13 to 50 (average 23 and standard deviation 6, *cf.* Figure 3.2). 56.13% of the subjects are male. Although there are a few family groups containing 3 and more siblings (Figure 3.1a), only one pair per group was used to avoid influence in the final classifier. There are no pair of subjects related by only one parent (*i. e.*, half-siblings) in the database and no investigation is performed for this special case, whatsoever. These numbers correspond to all images used in the

(a)

(b)

Figure 3.1: Examples of images of family groups belonging to HQ_{faces}: (a) three siblings, but the girl's image is discarded due to hair occlusion; and, (b) faces of three pairs of siblings in neutral frontal, smiling frontal and smiling profile poses.

experiments.

Several images had to be discarded due to some problems specified in the following. In spite of the highly controlled environment in which pictures were taken, there were some who either could not or should not be properly processed. This happened mainly due to hair occlusion, which in a few cases partially covered forehead but, most importantly, the eye, almost exclusively on profile poses. This compromised the detection of facial landmarks, as explained in Section 3.3. The latter occurred in cases of dense facial hair, which occluded facial properties as chin contour and skin color. When possible, such imperfections were manually corrected (re-positioning the landmarks) but a few images presenting one of both situations were discarded, decreasing the number pairs actually used in the simulations. For an example, refer to the rightmost image in Figure 3.1a.

The set of siblings pairs, after exclusion of improper images, was organized into three Individual Datasets (IDSs), as shown in Table 3.1, separating sets of subjects with common image poses:

- *HQ-f*: frontal expressionless images of 184 subjects (92 siblings' pairs);
- *HQ-fp*: 158 individuals, each represented by one frontal and one profile expressionless images (79 siblings' pairs);
- *HQ-fps*: 112 individuals, each represented by a set of four images per individual.

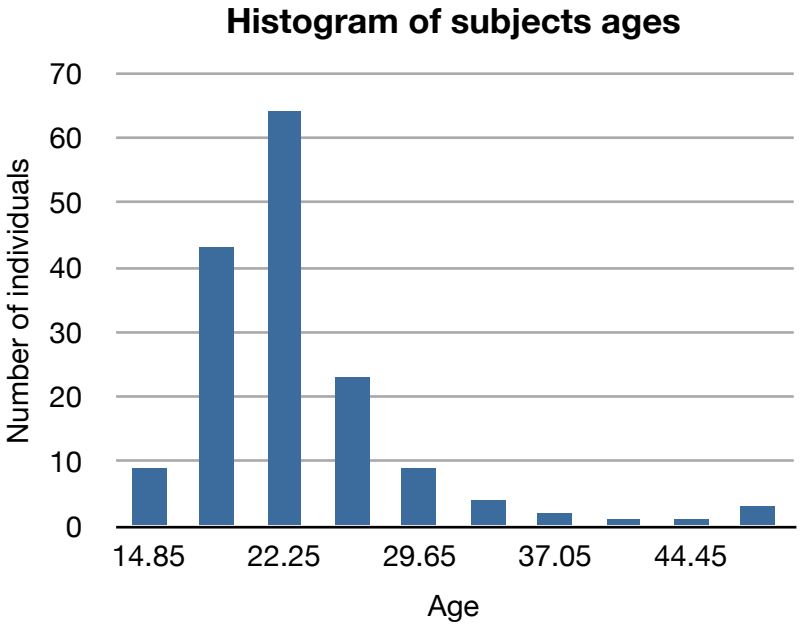


Figure 3.2: Distribution of ages of participants.

Two expressionless frontal and profile, and two smiling frontal and profile images (56 siblings' pairs).

Table 3.1: Summary of Individual Datasets (IDS).

Set	Subjects	Pairs	Pose			
			f	p	fs	ps
HQ-f	184	92	X			
HQ-fp	158	79	X	X		
HQ-fps	112	56	X	X	X	X

The mean age differences between siblings are shown in Table 3.2, for each dataset. This will be relevant in future discussions, where the correlation between the classification variable with the age differences will be evaluated (*cf.* Chapter 6).

3.2 Low Quality database

In order to evaluate the generalization capabilities of the proposed classification technique (see Chapter 5), a second database was prepared, composed of *low quality* images and, therefore, coined LQ_{faces}. It contains 196 individuals, totaling 98 pairs of siblings found over the Internet, where most of the subjects are celebrities. The photographs have different resolutions (approximately from few hundreds to more than 3,000 pixels across). The poses are semi-frontal. Faces often show expressions (smile), and images have been taken under different lighting conditions. Profiles and parent-child pairs are not available in LQ_{faces}. The individuals are 45.5% male, 87.9% Caucasian, 9.1% Afro-descendants and 3% Asiatic. Examples of siblings in LQ_{faces} are shown in Figure 3.3.

Both databases are available to the research community to foster further research on the siblings recognition problem.

Table 3.2: Averages and standard deviations of age differences between siblings who participated in the photo sessions.

	HQ-f	*HQ-fp*	*HQ-fps*
Age diff. (avg.±std.)	4.6 ± 4.6	4.7 ± 4.6	4.6 ± 5.2

Figure 3.3: Four pairs of siblings belonging to the LQ_{faces} image database.

3.3 Database normalization

In spite of the highly controlled environment in which a set of pictures is taken, they are likely to possess heterogeneity between them, especially those collected from the Internet. Each individual has its personal behavior in front of a camera. By leaning towards or away from it, and rotating the head, some people might influence the outcome of an automatic analysis by inserting non-uniformity throughout images, biasing the effects of facial features, especially those relying on geometrical distances (see Section 4.1). For these reasons, a pre-processing of all images is implemented, prior to the extraction of facial attributes.

3.3.1 Detection of facial landmarks

The very first step implemented is the detection of facial landmarks which are a set of points located onto key positions of the face. Such coordinates are the basis to all following procedures of the approach and any miss-detection can compromise further analysis.

For frontal images, 76 keypoints have been automatically computed with the Active Shape Models (ASMs) technique [21]. ASMs are statistical shape models that are fitted to an object in an image, combining both geometric and local appearance information. The statistical models are learned from a training set of labeled samples through an eigenspace-based approach. In this work, the open source implementation of ASMs available from [22, 23] has been used.

As for profile images, regardless of expression, 12 facial keypoints are identified with an algorithm derived from [24], based on the local curvature of the profile contour and on the local analysis of the image features.

3.3.2 Geometric normalization

After the detection of facial landmarks, the coordinates are used to align the images in the databases (DBs) and delimit the same area for all frontal and profile faces, including the most significant facial features. It is obtained by making coincident two facial landmarks with two reference points using suitable translation, rotation and isotropic scaling. These points are a subset of the landmarks, shown in Figure 3.4. The two landmarks used for geometric normalization of frontal images are the exterior eye corners (points 28 and 33 in Figure 3.4). The landmarks chosen for profile images are *nasion* (the point in the skull where the nasal and frontal bone unite) and *pogonion* (the most forward-projecting point on the anterior surface of the chin). These are, respectively, points 1 and 9 in Figure 3.4.

After geometric normalization, the face image is enclosed within a fixed rectangular area (*standard area*) with the two landmarks coincident with two predefined *reference positions* (Table 3.3). Observe that, in principle, this normalization could reduce the discriminative power of some of the geometric features used for siblings' classification, since absolute distance measurements could be relevant. However, absolute measures are not available for any images.

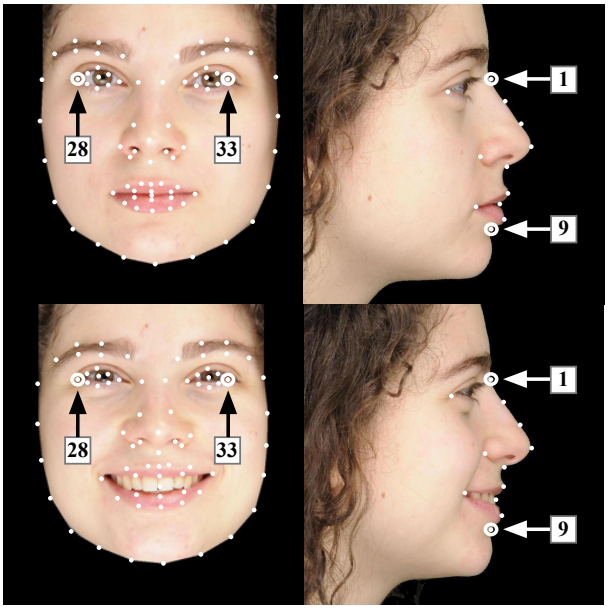


Figure 3.4: Landmarks detected on different image types. The numbers in the pictures pinpoint the elements chosen for geometric normalization.

Table 3.3: Reference positions used for geometric normalization. The image coordinate system has its origin on the top left and the x and y axes are, respectively, horizontal and vertical.

Image type	Standard area	Ref. 1 (x, y)	Ref. 2 (x, y)
Frontal (expressionless and smiling)	2000×2000	$(800, 800)$	$(1200, 800)$
Profile (expressionless and smiling)	2000×2000	$(1600, 800)$	$(1600, 1200)$

3.3.3 Background removal

Even after geometric normalization, the standard area background could also, potentially, affect some of the extracted features. For the images in HQ_{faces}, it has been subtracted using a simple chroma-keying technique, since volunteers' photos were taken placing them in front of a green screen. Segmentation has been performed manually for the LQ_{faces} set. For frontal images, the chin line defined by the ASM landmarks has been used to remove neck, shoulders and hairs as well.

3.3.4 Intensity normalization and face cropping

So far, the pre-processing stages to which all images are submitted are geometric normalization and background removal. Since images from HQ_{faces} have intensity normalized by the time of acquisition, this step is uniquely applied to images collected from the Internet. It is implemented by linearly mapping pixel values such that the minimum and maximum gray levels (on each channel separately) are mapped to 0 and 255, respectively.

There is yet another feature that requires another pre-processing task. Principal Component Analysis (PCA), also known as *eigenfaces*, is commonly applied to cropped faces, excluding most hair, background and even parts of the chin. Therefore, the keypoints were used to crop frontal and profile faces, yielding images such as those shown in Figure 3.5. The standard areas of frontal and profile images are presented in Table 3.4, which also includes the coordinates of the same reference points used for geometric normalization, previously discussed in Section 3.3.2.

Table 3.4: Characteristics of the reduced standard areas.

Image type	Standard area size	Ref. 1 (x,y)	Ref. 2 (x,y)
Frontal	600×600	(100,100)	(500,100)
Profile	450×600	(300,100)	(300,500)

Figure 3.5: Examples of normalized and cropped grayscale images of an individual in *HQ-fps*.

Chapter 4

Extraction of Features

In the previous chapter, the methodology used to assemble and pre-process two image databases (DBs) of siblings was discussed. The main DB, coined HQ_{faces}, consists of high quality photographs of siblings in different poses and expressions, taken from students and employees of the Politecnico di Torino. The second DB, LQ_{faces}, is composed by low-quality images of siblings collected from the Internet. The image processing approaches to normalize all images were also discussed, which aim at decreasing the heterogeneity among pictures, consequently reducing the influence of external factors into each person's facial characteristics.

In this chapter, the features extracted from all images are discussed. The general idea is to represent facial attributes as distances, textures and colors in terms of mathematical entities such as vectors and matrices. The facial characteristics and the mathematical values chosen to represent them are supposed to be highly descriptive, in order to provide a meaningful separation between classes. In other words, a collection of several (hopefully inherited) facial attributes is chosen, such that the differentiation between the value representing a pair of kins and the quantity describing a couple of unrelated people can be as easy as possible. The challenge here is finding a set of facial traits that could best represent the genetic sharing between siblings.

The facial characteristics shared by kins can be very different. For instance, although siblings, on average, have in common 50% of their genes, there are also extreme cases of siblings that do not seem to have any similarity between each other while conversely; there are identical twins, whose individuality in some cases cannot be perceived by simple inspection of their photographs [13]. The similarity between parent-child pairs is similar.

Supposedly, humans identify parent-child pairs by subjectively comparing phenotypic traits of each subject as, for instance, skin/ eyes/ hair colors, eyes, nose and mouth shapes, etc. Dal Martello and Maloney [14], on the basis of a high quality data set of children images, found that the upper part of the face carries more kinship clues.

Since it is not clear which are the exact facial traits able to effectively identify pairs of siblings, the strategy adopted in this work is to extract several features from each image. By considering facial characteristics of different nature, indeed is possible to find a set of attributes able to automatically identify kins, even more accurately than humans, as will be presented in Chapter 6.

Many different facial attributes were taken into consideration, and they can be divided into three main categories:

1. Geometric, focusing on geometrical positions and distances of facial landmarks.

2. Holistic, which considers the face as a whole, mixing geometric and texture information altogether.

3. Image descriptors, responsible for the representation of the texture around each face landmark.

For the sake of clarity, first consider Figure 4.1, which illustrates the feature extraction process of two subjects, a and b. Suppose that the subjects belong to set *HQ-fp*, which contains expressionless frontal and profile images, represented by subscripts f and p, respectively. First, K different facial attributes are extracted from each image and represented by matrices ϕ_k, $k = 1, \ldots, K$, for each individual. The n columns of a matrix ϕ are feature vectors $\mathbf{f} \in \Re^r$ of a given nature, as geometric distances or texture. The feature vector dimension r is independent on the pose, but the number of vectors extracted from the image is not. The comparison between two individuals is done per attribute, where a vector $\mathbf{v}_k^{(ab)}$ is computed by the n Euclidean distances of corresponding columns of matrices $\phi_k^{(a)}$ and $\phi_k^{(b)}$ extracted from respective subjects a and b, as illustrated in Figure 4.1. The final representative vector of a pair $\mathbf{x}^{(ab)}$ is the concatenation of differential vectors $\mathbf{v}_k^{(ab)}$ computed for each attribute k and each pose.

More formally, an Image Set $IS^{(a)}$ is a collection of images I_t representing one individual a in an Individual Dataset (IDS). The subscript t identifies the pose of the different images used to characterize an individual and it can assume the values f and p to denote,

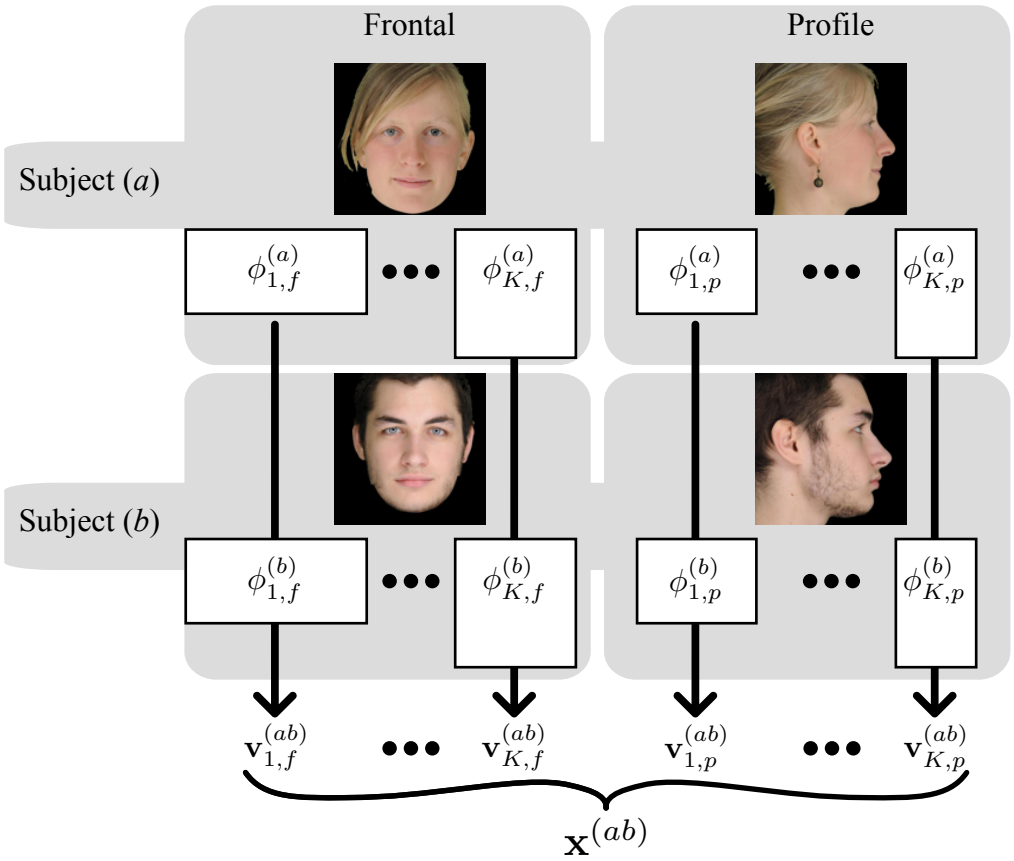


Figure 4.1: Feature extraction illustration.

respectively, *frontal* and *profile* expressionless images and fs and ps for *frontal* and *profile smiling* images, *i. e.*, $t \in \{f, fs, p, ps\}$.

An attribute $\phi_{k,t}$ of an image I_t is a collection of $n_{k,t}$ feature variables $\mathbf{f}_{k_t,j}$, $j = 1, \ldots, n_{k,t}$, where $\mathbf{f}_{k_t} \in \Re^{r_k}$. The subscript k indicates the type of features extracted from an image (*e. g.* the landmarks position or the lengths of their connecting segments); therefore, all $\mathbf{f}_{k_t}$ are obtained with the same technique and have the same dimension r_k. Each attribute can be represented by a matrix of size $r_k \times n_{k,t}$ as

$$\phi_{k,t} = \begin{pmatrix} \mathbf{f}_{k_t,1} & \mathbf{f}_{k_t,2} & \ldots & \mathbf{f}_{k_t,n_{k,t}} \end{pmatrix} \tag{4.1}$$

or

$$\phi_{k,t} = \begin{pmatrix} \mathbf{f}_{k_t,1}(1) & \mathbf{f}_{k_t,2}(1) & \ldots & \mathbf{f}_{k_t,n_{k,t}}(1) \\ \mathbf{f}_{k_t,1}(2) & \mathbf{f}_{k_t,2}(2) & \ldots & \mathbf{f}_{k_t,n_{k,t}}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{f}_{k_t,1}(r_k) & \mathbf{f}_{k_t,2}(r_k) & \ldots & \mathbf{f}_{k_t,n_{k,t}}(r_k) \end{pmatrix},$$

where each column of the matrix represents a feature vector. The attribute $\Phi_k^{(a)}$ of an individual a, is the concatenation of all attributes $\phi_{k,t}$ extracted from the images of different poses present in $IS^{(a)}$. For instance, the attributes Φ_k of the elements of *HQ-fp* are given by the matrix

$$\Phi_k = \begin{pmatrix} \phi_{k,f} & \phi_{k,p} \end{pmatrix}, \tag{4.2}$$

which has r_k rows and

$$\sum_{t \in \{f,p\}} n_{k,t}$$

columns. Similarly, for set *HQ-fps*, we have

$$\Phi_k = \begin{pmatrix} \phi_{k,f} & \phi_{k,fs} & \phi_{k,p} & \phi_{k,ps} \end{pmatrix}, \tag{4.3}$$

which also has r_k rows but

$$\sum_{t \in \{f,p,fs,ps\}} n_{k,t}$$

columns.

Finally, the characteristic vector $\mathbf{x}^{(ab)}$ for a couple of individuals a and b, is given by the vector of Euclidean distances, in their respective r-dimensional space, of the corre-

sponding feature variables composing $\phi_{k,t}^{(a)}$ and $\phi_{k,t}^{(b)}$, where subscript t is either f, $t \in \{f, p\}$ or $t \in \{f, fs, p, ps\}$. It can be easily seen that the characteristic vector of a couple is commutative, that is $\mathbf{x}^{(ab)} = \mathbf{x}^{(ba)}$. Figure 4.1 illustrates the process of computing vector $\mathbf{x}$ for a couple of individuals in *HQ-fp*, which contains frontal and profile poses.

In the following section the features extracted from the images are described. The mathematical formulation used to represent facial characteristics is introduced, and the rationale for choosing them are also presented. The basic idea is to model facial traits that are supposedly (or expectantly) associated with kinship relationship. According to genetic theory, the parents genotypes determine if some of their traits are more or less probable to be transferred to their children. Skin and eyes colors, as well as global and local facial geometric distances are some traits that might be considered by the human brain to determine whether two people are related or not. And, given the genetic sharing between siblings, one can expect that such traits can provide evidences to distinguish between siblings and non-siblings. For this reason, several features were extracted from each face, organized into *geometric*, *holistic* and *Local Image Descriptors*, as described next.

4.1 Geometric features

When asked which are the facial characteristics among siblings that are likely to be similar, one might suggest; different measurements among the face. Facial dimensions are, intuitively, inherited between kins. Therefore, many geometric dimensions were measured using the face landmarks detected using Active Shape Models (ASMs), as explained in Section 3.3. These dimensions are represented by (*1*) euclidean distances between corresponding landmarks of two subjects; (*2*) segment lengths uniting specific key points onto each face; (*3*) angles between those segments; and finally, (*4*) ratios between segments. These features are explained with more details in the following.

Attribute 1 (NPOS)

It contains the (x, y) normalized position of the facial landmarks. Since the two reference landmarks used for normalization have the same position for all the images of the same type (*cf.* Section 3.3), the attribute dimension is 2×74 for frontal and 2×10 for

profile images. For a pair, this attribute computes the distances between corresponding landmarks in the normalized images.

Attribute 2 (SEGS)

This attribute is computed from a dense net of segments, different for frontal and profile images, defined from the Delaunay Triangulation (DT) of the average position of the normalized landmarks over all the HQ_{faces} of the same type (frontal and profile). The lengths of these segments are effective in capturing global face shape and in describing, directly or through segment chains, representative facial measures (*e. g.* distances between mouth and eyes, mouth width, etc.). The number of segments is 184 for frontal and 25 for profile images. The reference net obtained from DT can be seen in Figure 4.2. With this attribute, each pair is characterized by the absolute differences between corresponding segment lengths.

Attribute 3 (ANGLES)

Angles of the triangles obtained from DT. In principle, angles can be computed from segment lengths and, therefore, they do not offer additional information. However, as we will show in the following, some feature variables can be discarded by a feature selection algorithm applied before the classification. As a result, the relationship between lengths and angles can be lost.

The angles are 342 for frontal images and 42 for profile images and the representative vector of a pair contains differences of corresponding angles.

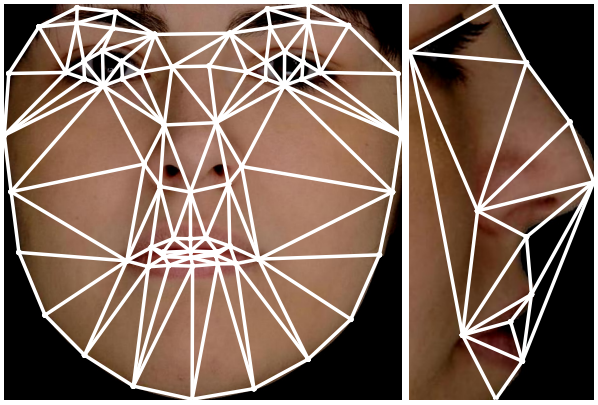


Figure 4.2: Reference segments, obtained from Delaunay Triangulation, for frontal (left) and profile (right) images.

Attribute 4 (DTR)

Delaunay Triangulation Ratios. This attribute contains the ratios between lengths of DT segments having a vertex in common. Two segments are used only once to compute a ratio, *i. e.*, once a ratio is computed, its inverse is not considered. The rationale of using DTR is that they are better suited than distances and angles to summarize (local) shape similarities, being invariant to isotropic and, to some extent, even anisotropic, scaling. For representing a frontal image, 862 ratios are considered, while for profiles, they are 92. For a pair, the attribute contains the set of differences between corresponding ratios.

4.2 Holistic attribute (PCA)

Attribute 5 (PCA)

Eigenfaces, first suggested by Sirovic and Kirby [25], have been extensively used for face image analysis in reduced dimensionality spaces. The main feature of the eigenfaces is that they capture both facial texture and geometry. Since it is not clear yet which are the facial elements more significant for detecting kinship clues for siblings, a first analysis using this popular catch-all technique for feature extraction was performed.

The general idea is to reduce the dimension of face images by considering their principal components, which is a set of eigenvectors (or *eigenfaces*) corresponding to the higher variance among the images being analyzed. Each original image is then projected onto this smaller dimensional space and is, therefore, represented by the vector containing the weights of a linear combination of the basis vectors.

The PCA was implemented on each pose set separately, to obtain the *eigenfaces* as those shown in Figure 4.3, that correspond to the highest eigenvalues. One can notice that the highest variance for frontal images is noticeable the chin contour, while other parts of the face, like the eyes, have smaller variance, mainly due to the geometric normalization performed as explained in Section 3.3. Similarly, the highest variance *eigenfaces* for profile images, regardless the expression, highlight the profile contour, indicating this is the feature with highest variance among different images.

Before applying Principal Component Analysis (PCA) to the set of available samples, the standard area of each individual's image was cropped to discard as many background pixels as possible, as previously explained in Section 3.3.4, to delimit the same section for

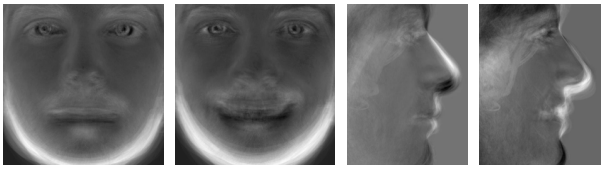


Figure 4.3: Examples of *eigenfaces* for set *HQ-fps*.

all frontal and profile faces, including only the most significant facial features (forehead, nose, mouth and chin; see Figure 3.5 for an example), and to remove as much as possible hairs in profile images. The size of these reduced regions and the position of the reference points inside them, are shown in Table 3.4 for different datasets. To deal with images taken in uncontrolled environments, *e. g.* those in LQ_{faces} and in PC_{faces}, we also applied contrast normalization.

4.3 Texture descriptors

Texture descriptors summarize the characteristics of image textures within regions. Several textural descriptors have been described in the literature. Since performing an exhaustive analysis of all of them is extremely difficult and out of the scope of this work, we focused on the ones that appear promising for characterizing our data. We recall that the image background in frontal and profile images does not influence descriptors in the face boundaries, since it has been removed during normalization.

4.3.1 Rotation-Invariant Co-occurrence of Local Binary Patterns (RIC-LBP)

The first texture descriptor is the Rotation Invariant Co-occurrence among Adjacent LBPs (RIC-LBP), proposed by Nosaka *et al.* [26]. In order to derive its formulation, a brief review of Local Binary Pattern (LBP) and Co-Occurrence of Adjacent LBPs (CoALBP) is provided.

Local Binary Patterns

Local Binary Pattern (LBP) [27] was originally designed as a texture description for a local region, called a micro-pattern, consisting of *binary patterns* that represent the magnitude relation between the center pixel of a local region and its neighboring pixels.

It is obtained by thresholding the image intensity of the surrounding pixels with that of the center pixel. To obtain an LBP histogram feature for use in classification, the binary patterns are converted to decimal numbers as labels, and then a histogram is generated from the labels of all local regions of an entire image.

More formally, the LBP at location $\mathbf{r} = (x \;\; y)^T$ of an image $I(\mathbf{r})$ is given by

$$\mathrm{LBP}(\mathbf{r}) = \sum_{i=0}^{N-1} \mathrm{sgn}\left[I(\mathbf{r} + \Delta\mathbf{s}_i) - I(\mathbf{r})\right] 2^i, \tag{4.4}$$

where $\Delta\mathbf{s}$ is the displacement vector from the center pixel to neighboring pixels given by

$$\Delta\mathbf{s}_i = \left(s\cos\theta_i, s\sin\theta_i\right),$$

$\theta_i = \dfrac{2\pi}{N} i;\; i = 0, \ldots, N-1$ and s is the scale parameter of the LBP.

The LBP computation around a pixel of gray intensity 120 located at $\mathbf{r}$ is illustrated in Figure 4.4 for $N = 8$. Specifically, the resulting binary pattern is 00100010_2 or, equivalently 34_{10}. The final representation of an image is given by the histogram of all 2^N possible labels of all patterns.

The main advantage of LBPs is its invariance to uniform changes in image intensity over an entire image, making it robust against changes in illumination. This is due to the fact that it considers only the magnitude relation between the center and neighboring pixel intensities. Owing to this characteristic, LBP has become a standard feature for texture and face recognition, and facial expression analysis.

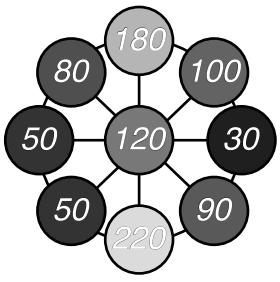


Figure 4.4: Illustration of how to compute a Local Binary Pattern (LBP). Resulting pattern is 00100010_2, or 34_{10}.

Co-Occurrence of Adjacent LBPs (CoALBP)

The original LBP does not preserve structural information among binary patterns, even though such information may be characteristic of an image. To tackle this limitation, Nosaka *et al.* [28] considered, in addition to LBPs, their co-occurrence in four given directions.

The CoALBP considers LBP pairs, *i. e.*,

$$\mathrm{P}(\mathbf{r}) = (\mathrm{LBP}(\mathbf{r}),\ \mathrm{LBP}(\mathbf{r} + \Delta\mathbf{r}))\,, \tag{4.5}$$

where

$$\Delta\mathbf{r} = (r\cos\theta, r\sin\theta)$$

and

$$\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}.$$

Parameter r specifies the distance between adjacent LBPs. Figure 4.5 illustrates how to compute the CoALBP feature from an image. Firstly, LBPs are computed onto the image using $N = 4$ (instead of typical $N = 8$) to reduce algorithm burden. Then, for each direction θ, the auto-correlation matrix for LBPs i and j are composed, generating the 3D bar graphs shown in Figure 4.5. And finally, these matrices are vectorized into a histogram, generating the final feature vector describing the entire image. The vector representation is in $\Re^{4N^2}$.

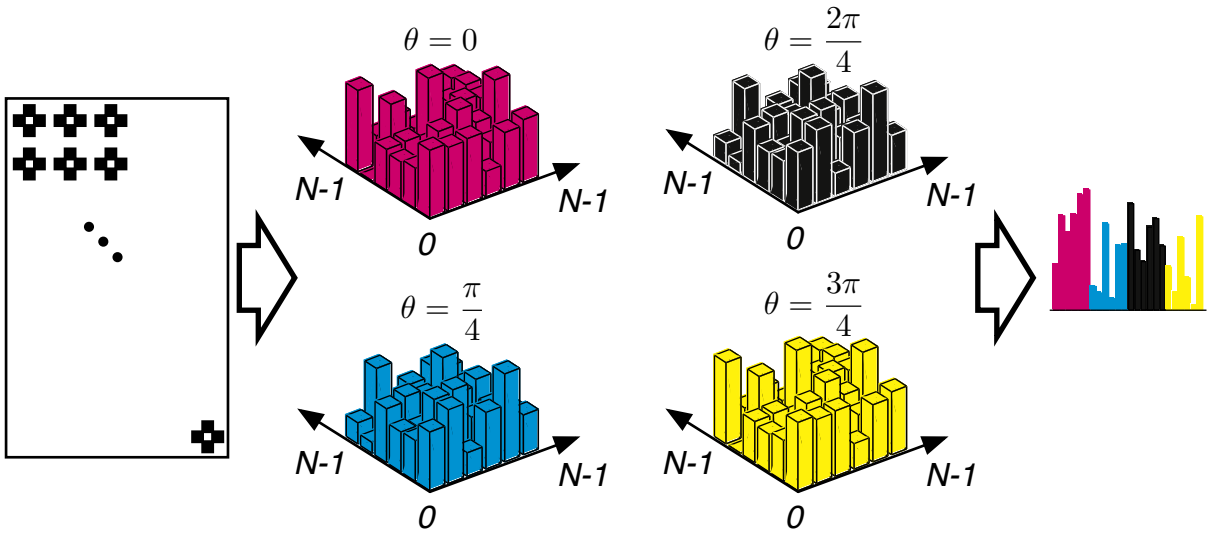


Figure 4.5: Illustration of how to compute CoALBP features from an image.

Attribute 6 (RIC-LBP)

Now, to derive the RIC-LBP, first modify the definition of LBP pairs to

$$\mathrm{P}_\theta(\mathbf{r}, \Delta\mathbf{r}_\theta) = [\mathrm{LBP}_\theta(\mathbf{r}), \mathrm{LBP}_\theta(\mathbf{r} + \Delta\mathbf{r}_\theta)]\,, \tag{4.6}$$

where

$$\mathrm{LBP}_\theta(\mathbf{r}) = \sum_{i=0}^{N-1} \mathrm{sgn}\left[I(\mathbf{r} + \Delta\mathbf{s}_{i,\theta}) - I(\mathbf{r})\right] 2^i \tag{4.7}$$

and

$$\Delta\mathbf{s}_{i,\theta} = [s\cos(\theta_i + \theta),\ \ s\sin(\theta_i + \theta)]\,. \tag{4.8}$$

where θ serves as the bias of the rotation angle in LBP. Based on the definition above, the LBP pair of each configuration has the same value in terms of rotation.

The final RIC-LBP histogram is built in such a way that adjacent LBPs being rotated by θ are summed up to the same bin, as illustrated in Figure 4.6. Firstly, adjacent LBPs rotated by $\theta = 0, \pi/4, \pi/2$ or $3\pi/4$ are labelled as equal. Then, those pairs rotated by 180° are analyzed. Finally, the final histogram is composed by summing the number of each CoALBP pair, regardless of direction, as depicted in Figure 4.6. This two-step analysis is performed to reduce the computational burden of pre-computing all possible CoALBP rotated pairs.

To choose between different LBP implementations, preliminary tests showed that the most effective for Kinship Verification (KV) is the RIC-LBP, whose implementation provides a 136 dimensional feature vector obtained from the whole grayscale image. The differences of corresponding elements are stored into the pair representative vector.

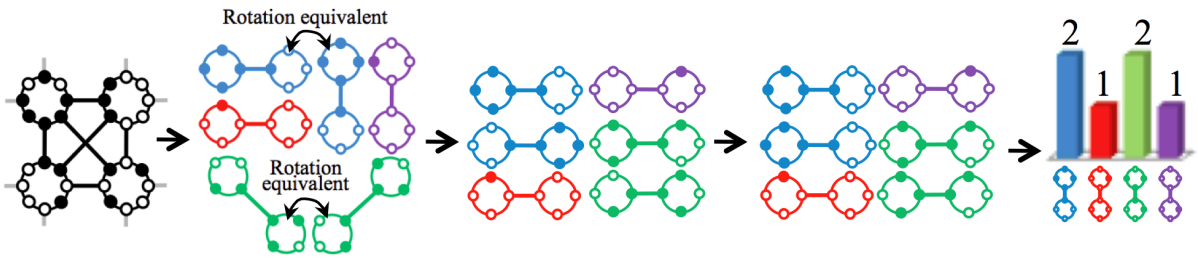


Figure 4.6: Illustration on how to compute the RIC-LBP feature.

4.3.2 Rotation-Invariant Gabor Feature (RIGF)

Attribute 7 (RIGF)

Gabor filters are widely used for edge detection and textures analysis. In our work we used the formulation of Rotation Invariant Gabor Feature (RIGF) proposed in [29] and illustrated in Figure 4.7, obtaining for each landmark a feature vector of dimension 96. It is computed on a window centered on each landmark, whose width was 15% of the distances between anchor points used for normalization (Section 3.3). This width was experimentally found to optimize the accuracy of the RIGF feature. The representative vector of a pair is obtained by computing the Euclidean distances between the RIGF vectors computed on corresponding landmarks of two subjects.

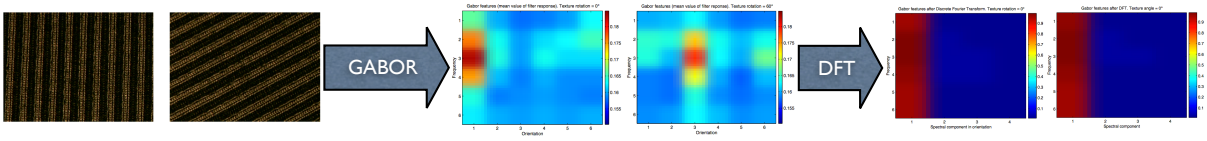


Figure 4.7: Example of RIGF feature computed onto an original and rotate texture.

4.3.3 Color descriptors

Attribute 8 - Local Color Descriptor (LCD)

Recently, different types of color-based descriptors have been introduced, in order to improve illumination invariance and discriminative power. According to literature [30], the descriptors applying Scale-Invariant Feature Transform (SIFT) [31] algorithm to different color schemes show better distinctiveness properties than the one based on other image characteristics, as histograms and color moments. The invariant properties related to changes in lighting that they exhibit, can be defined as follows. Given the unknown light source under which an image is taken, the changes of the illuminant in different images can be modeled as a combination of a scale factor applied to each of its (r, g, b) channels and a constant added to it. We have a *light intensity change* (1) when the image values are multiplied by the same factor; a *light shift* (2) when the same value is added to each channel; a *light intensity change and shift* (3) is a combination of the two; a *light color change* (4) is when each channel is multiplied by a different scale factor; a *light color change and shift* (5) is when, in addition to the previous condition, a different value is

added to each color channel. A descriptor is invariant to one of these properties when it does not depend on changes in the conditions expressed by the property.

In other words, considering a three channel rgb color image acquired under the influence of an unknown illuminant. The new $r'g'b'$ color space of a second image acquired under a changed illuminant can be modeled by a linear combination of the previous color space, i. e.,

$$\begin{cases} r' = \rho \cdot r + R \\ g' = \gamma \cdot g + G \\ b' = \beta \cdot b + B \end{cases} . \tag{4.9}$$

Depending on the scaling factors ρ, γ and β, and on the shifting constants R, G and B, different variations of colors can be defined, as outlined in Table 4.1.

We experimented with the following color descriptors:

- "*opponent*-SIFT", based on the opponent color space, having components o_1, o_2, o_3, derived from the RGB space as $o_1 = (r - g)/\sqrt{2}$, $o_2 = (r + g - 2b)/\sqrt{6}$ and $o_3 = (r + g + b)/\sqrt{3}$. *Opponent*-SIFT is invariant to properties 1, 2 and 3.

- "C-SIFT", based on a model of color invariance and reflected spectrum of colored bodies proposed in [32], which can be basically summarized by the normalized opponent colour space defined by the two components $o_1/o_3, o_2/o_3$. The third component is the intensity channel. C-SIFT in invariant to property 1.

- "*rg*-SIFT", encodes SIFT descriptors for the $\hat{r}$ and $\hat{g}$ chromaticity components of the normalized rgb color model, where $\hat{r} = r/(r + g + b)$ and $\hat{g} = g/(r + g + b)$. *rg*-SIFT is invariant to property 1.

- "*rgb*-SIFT" computes the SIFT descriptors on the RGB channels independently. It is invariant to all properties.

Table 4.1: Modeling of variations in light intensity and color.

Property	Light variation	Scaling factors	Shifting factors
1	Intensity change	$\rho = \gamma = \beta \neq 0$	$R = G = B = 0$
2	Intensity shift	$\rho = \gamma = \beta = 0$	$R = G = B \neq 0$
3	Intensity change and shift	$\rho = \gamma = \beta \neq 0$	$R = G = B \neq 0$
4	Color change	$\rho \neq \gamma \neq \beta$	$R = G = B = 0$
5	Color change and shift	$\rho \neq \gamma \neq \beta$	$R \neq G \neq B$

These color descriptors are computed evaluating a separate SIFT descriptor for each channel of the corresponding color models, whose size is 128 values, and concatenating them in a single vector. The similarity between two descriptors of the same type is represented by their Euclidean distance. With this attribute, a couple is represented by the vector of the distances of corresponding descriptors extracted from each individual's image set.

4.4 Features summary

All feature dimensions are summarized in Table 4.2, where the number of rows and columns composing each feature matrix are given for both attributes and IDS.

Table 4.2: Attributes dimensions.

	Attribute size (size of a feature × number of features)		Number of feature variables related to the attribute for an individual		
Attribute	Frontal (f) and frontal smiling (fs) images	Profile (p) and profile smiling (ps) images	*HQ-f*	*HQ-fp*	*HQ-fps*
1. NPOS	2×74	2×10	74	84	168
2. SEGS	1×184	1×25	184	209	418
3. ANGLES	1×342	1×42	342	384	768
4. DTR	1×862	1×92	862	954	1908
5. PCA	1×157 (f) 1×100 (fs)	1×129 (p) 1×93 (ps)	157	286	479
6. RIC-LBP	1×136	1×136	136	272	544
7. RIGF	96 × 76	96 × 12	76	88	176
8. LID	384×76	384×12	76	88	176

Chapter 5

Classification and Feature Selection

5.1 Support Vector Machines (SVM)

SVM are among the best "off-the-shelf" supervised learning algorithms. They were originally proposed by Vapnik [33] and improved by Vapnik and Corinna Cortes [34].

5.1.1 Intuition

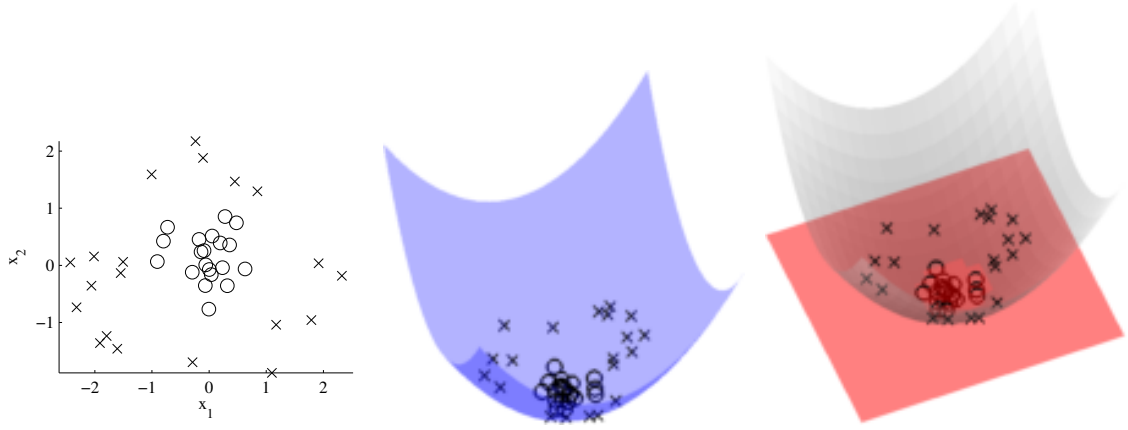
Prior to discussing SVM, consider the following example as a motivational intuition.

Consider a set of randomly generated sample vectors $\mathbf{x} \in \Re^2$ belonging to two classes represented by "points" and "circles", as shown in Figure 5.1a. It is noticeable that these two classes are not linearly separable on the $\Re^2$ plane. However, all samples $\mathbf{x} = \begin{pmatrix} x_1 & x_2 \end{pmatrix}^T$ shown in Figure 5.1a can be mapped into the three-dimensional space using the transformation

$$\mathbf{x}' = \begin{pmatrix} x_1 \\ x_2 \\ x_1^2 + x_2^2 \end{pmatrix}, \tag{5.1}$$

resulting in the plot shown in Figure 5.1b, where the samples representing class "circles" are located on the bottom of the surface. After the transformation, the data can be linearly separable by a plane in the three-dimensional space, as shown in Figure 5.1c. The higher the distance between the plane and the samples, the higher the confidence of the separation, *i. e.*, the better the classifier.

The hyperplane represents the classifier. A testing sample is firstly mapped onto the high dimensional space and its position with respect to the hyperplane, obtained from the



(a) (b) (c)

Figure 5.1: Intuition on how Support Vector Machines work: (a) randomly generated data samples of two different classes on 2D space; (b) samples mapped into a higher dimensional space (3D), and (c) linear separation of the data using a plane.

observation samples, is computed, defining to which class the new observation belongs to. The goal is finding both the mapping into a higher dimensional space and the separating plane that best separates the data. In the following a discussion is made on the concepts of functional and geometric margins, which relate the distance between samples and the separating plane.

5.1.2 Functional and geometric margins

More formally, consider the following mathematical formulation from [35], with a more general set of m linearly separable samples

$$\mathcal{D} = \left\{ \left(\mathbf{x}^{(i)}, y^{(i)} \right) \mid \mathbf{x}^{(i)} \in \Re^n, y^{(i)} \in \{-1, +1\} \right\}_{i=1}^{m}. \tag{5.2}$$

For the sake of simplicity, consider the particular case where $n = 2$, as shown in Figure 5.2. These samples can be separated by a hyperplane defined by all points $\mathbf{x}$ satisfying

$$\mathbf{w}^T \mathbf{x} + b = 0, \tag{5.3}$$

where parameters $\mathbf{w}$ and b are, respectively, a vector orthogonal to the separating hyperplane and a scalar with an absolute value equals to the distance between the hyperplane and the origin. Vector $\mathbf{w}$ is not unitary, since it is manipulated as will be explained

further. A sample $\mathbf{x}^{(i)}$ belongs to class "x" (crosses) if $\mathbf{w}^T\mathbf{x}^{(i)} + b \geq 0$; or to class "o" (circles) otherwise. Observe that the hyperplane divides $\Re^2$ in two subspaces, each of which corresponds to one class of the binary classification problem.

In general, a linear classifier for a binary classification problem with labels $y \in \{-1, 1\}$ and features $\mathbf{x} \in \Re^n$ is given by

$$h_{\mathbf{w},b}(\mathbf{x}) = g\left(\mathbf{w}^T\mathbf{x} + b\right), \tag{5.4}$$

where,

$$g(z) = \begin{cases} 1, \text{ if } z \geq 0 \\ -1, \text{ otherwise} \end{cases}. \tag{5.5}$$

Notice that, the classifier defined by Eq. (5.4) predicts either 1 or -1 based on a test sample $\mathbf{x}^{(i)}$ and parameters $\mathbf{w}$ and b, which are equivalent to classes "x" and "o" in Figure 5.2, respectively.

Given one training example $\left(\mathbf{x}^{(i)}, y^{(i)}\right)$, we define the **functional margin** of $(\mathbf{w}, b)$ with respect to the training example i as follows

$$\hat{\gamma}^{(i)} \equiv y^{(i)}\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right). \tag{5.6}$$

Notice that if $y^{(i)} = 1$, then for the functional margin to be large (*i. e.*, for the prediction to be confident and correct), $\mathbf{w}^T\mathbf{x} + b$ should be a large positive number. Conversely, if $y^{(i)} = -1$, then for the functional margin to be large, $\mathbf{w}^T\mathbf{x} + b$ should be a large negative

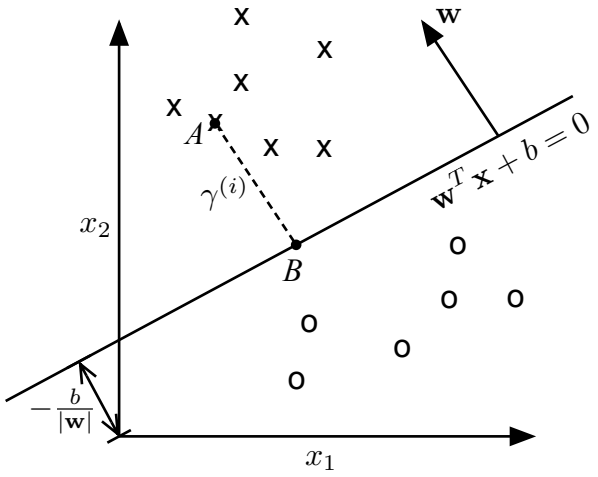


Figure 5.2: Separation of two classes by a hyperplane defined by parameters $\mathbf{w}, b$.

number. Moreover, if $y^{(i)}\left(\mathbf{w}^T\mathbf{x}+b\right) > 0$, the prediction on this example is correct. Hence, a large functional margin represents a confident and a correct prediction. The function margin of $(\mathbf{w}, b)$ with respect to $\mathcal{D}$ is the smallest of the functional margins of the individual training samples, *i. e.*,

$$\hat{\gamma} \equiv \min_{i=1,\ldots,m} \hat{\gamma}^{(i)}. \tag{5.7}$$

Similarly, the geometric margin with respect to one sample is defined as the distance $\gamma^{(i)}$ between positive sample $\mathbf{x}^{(i)}$ (point A) and the decision boundary (point B), as shown in Figure 5.2. Point B is given by

$$\mathbf{x}^{(i)} - \gamma^{(i)}\mathbf{w}/|\mathbf{w}|$$

since it lies on the decision boundary (Eq. 5.3). Hence, we have that

$$\mathbf{w}^T\left(\mathbf{x}^{(i)} - \gamma^{(i)}\frac{\mathbf{w}}{|\mathbf{w}|}\right) + b = 0,$$

or

$$\gamma^{(i)} = \left(\frac{\mathbf{w}}{|\mathbf{w}|}\right)^T \mathbf{x}^{(i)} + \frac{b}{|\mathbf{w}|}.$$

More generally, the geometric margin of $(\mathbf{w}, b)$ with respect to a training example i (positive or negative) is

$$\gamma^{(i)} \equiv y^{(i)}\left(\left(\frac{\mathbf{w}}{|\mathbf{w}|}\right)^T \mathbf{x}^{(i)} + \frac{b}{|\mathbf{w}|}\right) \tag{5.8}$$

Finally, the geometric margin with respect to the entire training set $\mathcal{D}$ is

$$\gamma \equiv \min_{i=1,\ldots,m} \gamma^{(i)}. \tag{5.9}$$

Notice that the functional and the geometric margins are related by

$$\gamma = \frac{\hat{\gamma}}{|\mathbf{w}|}. \tag{5.10}$$

Moreover, by scaling parameters $|\mathbf{w}|$ and b, the functional margin can be made arbitrarily large, given that $g(z)$ from Eq. (5.5) does not change by scaling z, and consequently, the classifier $h_{\mathbf{w},b}(\mathbf{x})$ defined in Eq. (5.4) remains unmodified. On the other hand, the

geometric margin do not change with scaling of the parameters, since it is normalized by $|\mathbf{w}|$. Because of this invariance to scaling the parameters, when trying to fit $\mathbf{w}$ and b to the training data, an arbitrary constraint on $\mathbf{w}$ can be imposed without changing the actual classifier; for instance, parameters can be forced to satisfy $|\mathbf{w}| = 1$, or $|w_1| = 4$, or $w_1^2 + |w_2| = 17$. In conclusion, the idea is scaling parameters $\mathbf{w}$ and b such as to maximize the functional margin without changing the actual classifier.

5.1.3 The optimal margin classifier

At first, considering linearly separable data, the objective of trying to find a decision boundary that maximizes the geometric margin might seem the main goal, based on the fact that this would result in confident predictions (a good fit to the training data). This classifier would then properly separate positive and negative samples with a "gap" given by the geometric margin.

Assuming that it is possible to separate all positive and negative training samples with some separating hyperplane, the following optimization problem can be posed:

$$\begin{aligned} \max_{\gamma,\mathbf{w},b} \quad & \gamma \\ \text{s. t.} \quad & \begin{cases} y^{(i)}\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right) \geq \gamma \\ i = 1, \ldots, m \\ |\mathbf{w}| = 1 \end{cases} \end{aligned} , \tag{5.11}$$

which states that the objective is to maximize γ subject to each training example having functional margin at least γ. In addition, the $|\mathbf{w}| = 1$ constraint guarantees that the functional margin equals the geometric margin, ensuring that the geometric margins are also at least γ. Ultimately, solving this problem fits $\mathbf{w}, b$ to the training data with the largest possible geometric margin with respect to the training set.

However, the imposition $|\mathbf{w}| = 1$ is a non-convex constraint, making the problem difficult to solve using standard optimization software. The problem stated in (5.11) can be modified to

$$\begin{aligned} \max_{\gamma,\mathbf{w},b} \quad & \frac{\hat{\gamma}}{|\mathbf{w}|} \\ \text{s. t.} \quad & \begin{cases} y^{(i)}\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right) \geq \hat{\gamma} \\ i = 1, \ldots, m \end{cases} \end{aligned} , \tag{5.12}$$

which means maximizing $\hat{\gamma}/|\mathbf{w}|$, subject to the functional margins all being at least $\hat{\gamma}$. The $|\mathbf{w}|$ constraint was dropped and, since the geometric and functional margins are related by Eq. (5.10), the problem stated in (5.12) gives the optimal geometric margin. However, the objective function $\hat{\gamma}/|\mathbf{w}|$ still relates to a non-convex problem.

But, as discussed previously the final classifier does not change by arbitrarily scaling parameters $\mathbf{w}$ and b (recall that the functional margin can be made arbitrarily large by such scaling while the geometric margin stays the same). A new constraint on the functional margin of $\mathbf{w},\ b$ can be imposed

$$\hat{\gamma} = 1, \tag{5.13}$$

which can be satisfied by scaling $\mathbf{w}$ and b. Plugging (5.13) into (5.12), and noting that maximizing $\hat{\gamma}/|\mathbf{w}|$ is the same as minimizing $|\mathbf{w}|^2$, the following optimization problem is derived

$$\begin{array}{ll} \min_{\gamma,\mathbf{w},b} & \frac{1}{2}|\mathbf{w}|^2 \\ \text{s. t.} & \begin{cases} y^{(i)}\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right) \geq 1 \\ i = 1, \ldots, m \end{cases} . \end{array} \tag{5.14}$$

The optimization problem (5.14) has a convex quadratic objective function and linear constraints, whose solution can be found using standard quadratic programming (QP) code, which provides the Optimal Margin Classifier (OMC).

Although a standard QP code could provide a solution to problem (5.14) in the following, a better solution is presented, utilizing the Lagrange duality. This will allow the use of kernels to efficiently solve our optimization problem in high dimensional spaces.

5.1.4 Lagrange duality

Problem (5.14) can be modified in such a way to permit the use of kernels, which allows an efficient algorithmic solution for samples belonging to very high dimension. In order to derive this alternative, let's first consider a constrained optimization problem given by

$$\begin{array}{ll} \min_{\mathbf{w}} & f(\mathbf{w}) \\ \text{s. t.} & \begin{cases} g_i(\mathbf{w}) = 0 \\ i = 1, \ldots, m \end{cases} , \end{array} \tag{5.15}$$

which involves minimizing a function f subject to constraints given by $g_i = 0$ for $i = 1, \ldots, m$. To solve this optimization problem, one can use Lagrange Multipliers (LM), a strategy for finding local maxima/minima of a function subject to equality constraints [36]. It consists in defining an auxiliary function called the *Lagrangian*

$$\mathcal{L}(\mathbf{w}, \beta_i) = f(\mathbf{w}) + \sum_{i=1}^{l} \beta_i g_i(\mathbf{w}), \tag{5.16}$$

and finding the solution by making

$$\nabla_{\mathbf{w},\beta_i} \mathcal{L}\left(\mathbf{w}, \beta_i\right) = 0. \tag{5.17}$$

The β_i's are called the *Lagrange Multipliers*.

Now, consider the following, called the **primal** optimization problem:

$$\begin{array}{ll} \min_{\mathbf{w}} & f(\mathbf{w}) \\ \text{s. t.} & \begin{cases} g_i(\mathbf{w}) \leq 0, \ i = 1, \ldots, k \\ h_i(\mathbf{w}) = 0, \ i = 1, \ldots, l \end{cases} \end{array} . \tag{5.18}$$

Which can be solved using the *generalized Lagrangian*

$$\mathcal{L}(\mathbf{w}, \alpha, \beta) = f(\mathbf{w}) + \sum_{i=1}^{k} \alpha_i g_i(\mathbf{w}) + \sum_{i=1}^{l} \beta_i h_i(\mathbf{w}),$$

where the α_i's and β_i's are the Lagrange multipliers. Consider the quantity

$$\theta_{\mathcal{P}}(\mathbf{w}) = \max_{\alpha,\beta:\alpha_i \geq 0} \mathcal{L}(\mathbf{w}, \alpha, \beta).$$

Where subscript $\mathcal{P}$ stands for *primal*. If a given $\mathbf{w}$ violates any of the primal constraints (*i. e.*, if either $g_i(\mathbf{w}) > 0$ or $h_i(\mathbf{w}) \neq 0$ for some i), then

$$\theta_{\mathcal{P}} = \infty.$$

And if a given $\mathbf{w}$ satisfies the constraints, then $\theta_{\mathcal{P}} = f(\mathbf{w})$, *i. e.*,

$$\theta_{\mathcal{P}}(\mathbf{w}) = \begin{cases} f(\mathbf{w}) & \text{if } \mathbf{w} \text{ satisfies primal constraints} \\ \infty & \text{otherwise} \end{cases} . \tag{5.19}$$

So, for all values of $\mathbf{w}$ satisfying the primal constraints, the optimization problem (5.18) can be rewritten as

$$\min_{\mathbf{w}} \theta_{\mathcal{P}}(\mathbf{w}) = \min_{\mathbf{w}} \max_{\alpha,\beta:\alpha_i \geq 0} \mathcal{L}(\mathbf{w}, \alpha, \beta), \tag{5.20}$$

whose objective is by $p^* = \min_{\mathbf{w}} \theta_{\mathcal{P}}(\mathbf{w})$; called the *value* of the primal problem.

Observe that in the primal optimization problem, the objective is to maximize the Lagrangian with respect to α, β. Now consider a similar optimization problem, called the *dual* problem

$$\max_{\alpha,\beta:\alpha_i \geq 0} \theta_{\mathcal{D}}(\alpha, \beta) = \max_{\alpha,\beta:\alpha_i \geq 0} \min_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \alpha, \beta). \tag{5.21}$$

Which is exactly the same as the primal problem, except for the exchange of the "max" and "min". Also, let the value of the dual problem be given by $d^*_{\alpha,\beta:\alpha_i \geq 0} \theta_{\mathcal{D}}(\mathbf{w})$.

The primal and dual optimization problems are related by the observation that the "max min" of a function is always less or equal than its "min max";

$$d^* = \max_{\alpha,\beta:\alpha_i \geq 0} \min_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \alpha, \beta) \leq \min_{\mathbf{w}} \max_{\alpha,\beta:\alpha_i \geq 0} \mathcal{L}(\mathbf{w}, \alpha, \beta) = p^* \tag{5.22}$$

But, under certain conditions,

$$d^* = p^*, \tag{5.23}$$

so that the dual problem can be solved in order to find a solution to the primal problem. In the following these conditions are presented.

Suppose f and each g_i are convex, and each h_i is affine. Suppose further that the constraints on g_i are strictly feasible. This means that there exists some $\mathbf{w}$ so that $g_i(\mathbf{w}) < 0 \ \forall \ i$. Under these assumptions, there must exist $\mathbf{w}^*, \alpha^*, \beta^*$ so that $\mathbf{w}^*$ is the solution to the primal problem, α^*, β^* are the solution to the dual problem, and moreover

$$p^* = d^* = \mathcal{L}(\mathbf{w}^*, \alpha^*, \beta^*). \tag{5.24}$$

In addition, $\mathbf{w}^*, \alpha^*$ and β^* satisfy the Karush-Kuhn-Tucker (KKT) conditions:

$$\frac{\partial}{\partial \mathbf{w}_i}\mathcal{L}(\mathbf{w}^*, \alpha^*, \beta^*) = 0,\ i = 1, \ldots, n, \tag{5.25a}$$

$$\frac{\partial}{\partial \beta_i}\mathcal{L}(\mathbf{w}^*, \alpha^*, \beta^*) = 0,\ i = 1, \ldots, l, \tag{5.25b}$$

$$\alpha_i^* g_i(\mathbf{w}^*) = 0,\ i = 1, \ldots, k, \tag{5.25c}$$

$$g_i(\mathbf{w}^*) \leq 0,\ i = 1, \ldots, k, \tag{5.25d}$$

$$\alpha^* \geq 0,\ i = 1, \ldots, k. \tag{5.25e}$$

Moreover, if some $\mathbf{w}^*$, α^*, β^* satisfy the KKT conditions, then it is also a solution to the primal and dual problems.

Equation 5.25d is called the KKT dual complementarity condition. Specifically, it implies that if $\alpha^* > 0$, then $g_i(\mathbf{w}^*) = 0$. This means that the constraint $g_i(\mathbf{w}) \leq 0$ is active, meaning that it holds with equality rather than with inequality. Later on, this will be the key for showing that the SVM has only a small number of "support vectors"; the KKT dual complementarity condition will also provide a test for convergence when the Sequential Minimal Optimization (SMO) algorithm will be discussed.

5.1.5 Optimal margin classifiers

Recall the primal optimization problem for finding the optimal margin classifier given by Eq. (5.14)

$$\begin{aligned} &\min_{\gamma,\mathbf{w},b} \quad \tfrac{1}{2}|\mathbf{w}|^2 \\ &\text{s. t.} \quad \begin{cases} y^{(i)}\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right) \geq 1 \\ i = 1, \ldots, m \end{cases} . \end{aligned}$$

Since there is no equality constraint, the Lagrangian has only "α_i's" terms:

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2}|\mathbf{w}|^2 - \sum_{i=1}^{m} \alpha_i \left[y^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b) - 1\right]. \tag{5.26}$$

In order to obtain the dual of this problem, we minimize the Lagrangian with respect to $\mathbf{w}$ and b for a given α,

$$\nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w}, b, \alpha) = \mathbf{w} - \sum_{i=1}^{m} \alpha_i y^{(i)}\mathbf{x}^{(i)} = 0,$$

which implies that

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i y^{(i)} \mathbf{x}^{(i)}. \tag{5.27}$$

The derivative with respect to b yields

$$\frac{\partial}{\partial b}\mathcal{L}(\mathbf{w}, b, \alpha) = \sum_{i=1}^{m} \alpha_i y^{(i)} = 0. \tag{5.28}$$

Plugging the definition of $\mathbf{w}$ (Eq. 5.27) into the Lagrangian (Eq. 5.26), and considering Eq. (5.28), yields

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y^{(i)} y^{(j)} \alpha_i \alpha_j \left\langle \mathbf{x}^{(i)}, \mathbf{x}^{(i)} \right\rangle. \tag{5.29}$$

Equation (5.29) was obtained by minimizing the Lagrangian with respect to $\mathbf{w}$ and b. Now, to derive the dual optimization problem, the following step is to maximize the same formulation with respect to α, considering the constrains $\alpha_i \geq 0$ and Eq. (5.28), *i. e.*,

$$\begin{gathered} \max_{\alpha} \; W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y^{(i)} y^{(j)} \alpha_i \alpha_j \left\langle \mathbf{x}^{(i)}, \mathbf{x}^{(i)} \right\rangle, \\ \text{s. t. } \alpha_i \geq 0, \; i = 1, \ldots, m, \\ \sum_{i=1}^{m} \alpha_i y^{(i)} = 0. \end{gathered} \tag{5.30}$$

One can verify that the KKT conditions are indeed satisfied for this optimization problem. Hence, one can solve the dual in lieu of solving the primal problem. Specifically, in the dual problem above, we have a maximization problem in which the parameters are the $\alpha_i's$. If the problem can be solved, *i. e.*, one can find the $\alpha_i's$ that maximize $W(\alpha)$ subject to the constraints, then Eq. (5.27) can be used to find the optimal $\mathbf{w}'s$ as a function of the $\alpha_i's$. Having found $\mathbf{w}^*$, by considering the primal problem, it is also straightforward to find the optimal value for the intercept term b as

$$b^* = -\frac{1}{2}\left(\max_{i:y^{(i)}=-1} \mathbf{w}^{*T} \mathbf{x}^{(i)} + \min_{i:y^{(i)}=1} \mathbf{w}^{*T} \mathbf{x}^{(i)} \right). \tag{5.31}$$

Taking a closer look at Eq. (5.27), which provides the optimal value of $\mathbf{w}$ in terms of (the optimal value of) α. Supposing the model's parameters were fit to a training set, and now a prediction of a new input $\mathbf{x}$ must be made. In order to do that, one calculates $\mathbf{w}^T\mathbf{x} + b$ and predicts $y = 1$ if and only if this quantity is bigger than zero. But, by

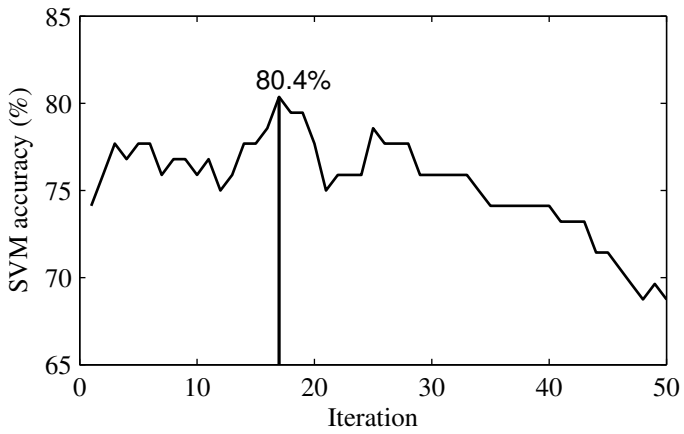


Figure 5.3: Sequential Forward Selection strategy: an example (for C-SIFT attribute and *HQ-fps* dataset) of the behavior of the classification accuracy on different iterations.

using (5.27), this quantity can also be rewritten:

$$\mathbf{w}^T\mathbf{x} + b = \left(\sum_{i=1}^{m} \alpha_i y^{(i)} \mathbf{x}^{(i)}\right)^T \mathbf{x} + b \tag{5.32}$$

$$= \sum_{i=1}^{m} \alpha_i y^{(i)} \left\langle \mathbf{x}^{(i)}, \mathbf{x} \right\rangle + b. \tag{5.33}$$

Hence, if each α_i is computed, in order to make a prediction, one must calculate a quantity that depends only on the inner product between $\mathbf{x}$ and the points in the training set.

By examining the dual form of the optimization problem, an entire algorithm in terms of only inner products between feature vectors was derived. In the next section, this property is exploited to apply kernels to this classification problem. The resulting algorithm is able to efficiently learn in very high dimensional spaces.

5.2 Feature Selection

Feature Selection (FS) is a data preprocessing step that is frequently applied in machine learning. FS extracts the subset of features used to describe the data that is optimal under certain criteria. For classification problems, the criterion to meet is the improvement of the classification accuracy. The purposes of FS are reducing the dimensionality of the input data, removing irrelevant information and improving the comprehension of data and results by telling which are the most important features and how they are correlated.

Despite the fact that SVMs are generally acknowledged for their generalization capa-

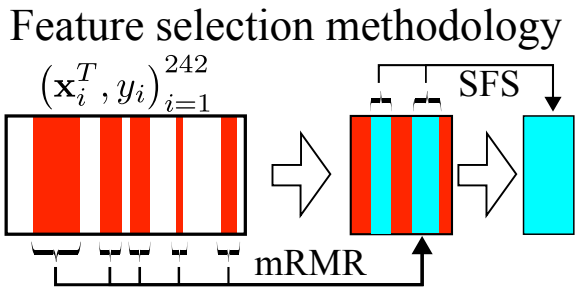


Figure 5.4: Illustration of the Feature Selection (FS) procedure.

bilities, their integration with FS schemes provides several improvements [37, 38]. Thus, in this work the following robust two-step feature selection methods were applied.

First step (mRMR). It is based on min-Redundancy Max-Relevance (mRMR) algorithm, whose better performance over the conventional top-ranking methods has been widely demonstrated in literature [39]. The mRMR algorithm sorts the features that are most relevant for characterizing the classification variable, pointing at the contemporaneous minimization of their mutual similarity and maximization of their correlation with the classification variable. The number of the candidate features selected by mRMR was heuristically set to 50 for each dataset and each characterizing attribute.

Second step (SFS). The output of mRMR is a generic candidate feature set not necessarily optimal for the chosen classifier. Therefore, we applied, as second FS step, a Sequential Forward Selection (SFS) scheme [40]. The SFS scheme is widely used for its simplicity. Starting from an initially empty feature set S, SFS adds, at each iteration, the feature providing the greatest improvement of the classification accuracy until no more improvements can be obtained. Since this stopping criterion tends to trap the algorithm in local minima, in our approach, we proceeded with the iterations until all features were added to S, and then we selected the feature set corresponding to the iteration that obtained the best classification accuracy (Figure 5.3).

The two-step FS process is illustrated in Figure 5.4. The matrix composed for each dataset is illustrated on the left. Each of its rows is a representative vector of a couple $\mathbf{v}^{(a,b)}$ computed as illustrated in Figure 4.1. The columns are then firstly selected by the mRMR algorithm. Then, the SFS algorithm is applied to the chosen features, yielding the final set composed of the "best-of-the-best" features.

Finally, the entire algorithm is illustrated in Figure 5.5. It is based on Dissimilarity-Based Classifications (DBCs) [41]. The first step consists in normalizing all images from

all Pair Datasets (PDSs). Then, features of different nature are extracted, so that each image is converted to a mathematical representation. Then, these representations are combined to compute the representative vectors for each sample, *i. e.*, a pair of siblings (positive) or non-siblings (negative). Finally, a two-step feature selection scheme, associated with SVM classification is applied onto these datasets to extract the characteristics that show more descriptive capabilities related to the siblings classification problem.

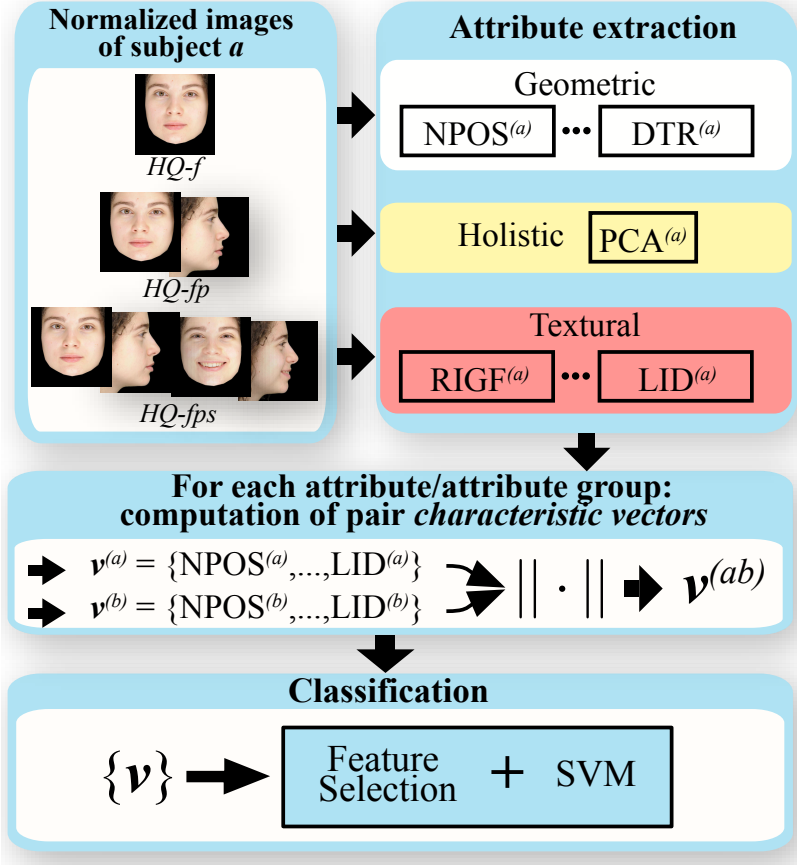


Figure 5.5: Outline of the sibling recognition process.

Chapter 6

Results and Discussion

This chapter presents the results obtained during the development of this thesis. Given the heuristic nature of the proposed investigation, *i. e.*, considering that the concept of similarity of faces is much more encompassing than the concept of identity, several simulations were executed using different configurations, parameters, machines, platforms, programming languages, etc, as well as software developed by third parties. As discussed previously, the kinship recognition problem has been proposed recently in the pattern analysis/ machine learning field. Due to the lack of touchstones available at the beginning of the investigation, a comprehensive number of experiments attempting to effectively identifying siblings was performed. Although an extensive discussion of the details of all experiments is not provided here for brevity, some relevant observations and results are presented. In spite of the fact that these initial observations did no provide good numeric results, they guided the research through different paths, ultimately providing relevant results.

A brief overview of unsatisfactory results is presented in section 6.1. Sections 6.2 through Section 6.5 discuss the relevant results published in [5, 6, 7].

6.1 Preliminary observations

This section describes the first steps that matured the investigation toward the effective discrimination between siblings and non-siblings. Although not significant in accuracy, the observations were important in guiding the research towards the final relevant results presented in Section 6.2 through Section 6.5.

6.1.1 Analysis of facial distances

The very first attempts to classify siblings were performed considering geometric distances obtained from the faces. This is an intuition of facial similarities. In other words, since kins are likely to share phenotype traits, relative distances throughout the face could be considered to contain kinship information. Several initial facial distances were tested, such as the distances between:

- left and right eyes;

- nose tip and mouth;

- mouth to chin tip;

- eyes to nose tip; etc.

In addition, several geometric properties of the face were tested, such as:

- nose length;

- length between the lateral alas of the nose;

- mouth width;

- eyes, eyebrows and mouth circumscribing perimeters;

- perimeter of chin contour;

- relations between facial segments, such as the ratio between horizontal and vertical axes of ellipses fitted to eyes, mouth, eyebrows and nose landmarks, etc.

However, such punctual geometric distances provided unsatisfactory results, achieving a maximum of ≈65% accuracy. When applying the classification algorithm based on SVM with grid optimization of the SVM parameters γ and C (*cf.*, Section 6.5), for instance, the accuracy optimization is shown in Figure 6.1 when the feature being extracted is the distance between nose tip and mouth center. Notice that an accuracy of 63% is achieved, which gives some indication that distances can be used for the classification. However, a more comprehensive and systematic scheme for the extraction of geometric features should be considered.

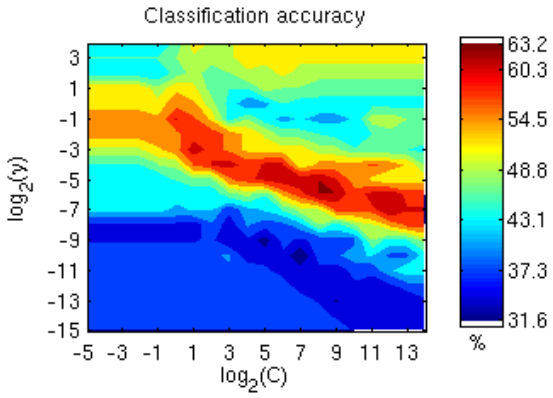


Figure 6.1: Optimization of SVM parameters yielding a 63.2% accuracy in siblings classification considering only the euclidean distance between nose tip and mouth.

These first observations of the classification performance using geometric distances led the investigation to the use of Delaunay Triangulation (DT) for the extraction of a considerable dense net of segments from the faces. Specifically, the use of ratios between Delaunay Triangulation segments (*cf.*, Section 4.1) was motivated by these observations. Indeed, ratios between facial segments are considerably effective in the classification of siblings (*cf.*, Section 6.5) and parent-child pairs (*cf.*, Section 7.3).

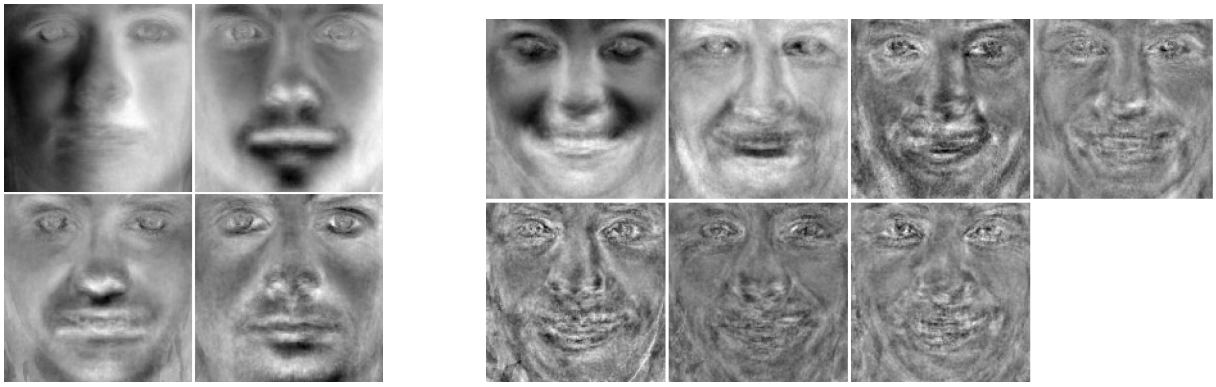
6.1.2 Siblings classification using only *eigenfaces*

Another question asked in the beginning of the research was if siblings could be discriminated using only their holistic representation. To answer this question, the simplest way of extracting holistic information from the faces, namely, Principal Component Analysis (PCA), was used (*cf.* Section 4.2). The rationale was that:

- given the simplicity of the PCA method, any other accuracies obtained with different characteristics should outperform those provided by PCA, and;

- by describing each face as a linear combination of the *eigenfaces* associated with the highest variance of the data, some facial characteristics related to the siblings classification problem could be highlighted by observing the *eigenfaces* surviving the FS pruning.

Figure 6.2 shows some *eigenfaces* associated with the highest *eigenvalues* in decreasing order of importance from left to right and top to bottom for the expressionless frontal



(a) **(b)**

Figure 6.2: Most relevant *eigenfaces* for the siblings classification problem in decreasing order of importance from left to right and top to bottom: (a) best *eigenfaces* for *HQ-f* PDS, and; (b) best *eigenfaces* for LQ_{faces} dataset.

high quality Pair Dataset (PDS) *HQ-f* (Figure 6.2a) and the low quality dataset LQ_{faces} (Figure 6.2b) after the implementation of the classification and feature selection algorithm.

6.1.3 Scale-Invariant Feature Transform (SIFT)

Another investigation used the Scale-Invariant Feature Transform (SIFT) [31] and the Speeded Up Robust Features (SURF) [42] onto grayscale images. The latter has similar descriptive capabilities in comparison with the former, but is much faster to compute. They are aimed at detecting invariant keypoints from an image and describing each keypoint by a vector for posterior matching. They are widely used for object detection and recognition. Our implementation considers only the description of image windows centered on each facial landmark obtained onto the face as explained in Section 3.3.1, *i. e.*, we did not detect invariant keypoints from SIFT and SURF algorithms. Rather, we used only their techniques in describing each window around each landmark. SIFT provides a 128-dimension feature vector for each of the 76 frontal and 12 profile keypoints. On the other hand, SURF provides a 64-dimension feature vector for each of the facial landmarks.

Both SIFT and SURF algorithms are patented and this work makes no claim about their copyrights whatsoever. Their open-source implementation can be found in version 2.1 of the Open-Source Computer Vision library (OpenCV) [43] for non-commercial purposes.

One first observation was the descriptive superiority when using descriptors computed around each landmark using SIFT and SURF in comparison with punctual geometric distances as discussed previously in Section 6.1.1. For instance, SIFT achieved a classification accuracy of 71% and 65% for expressionless frontal and profile datasets, as can be appreciated in Figure 6.3. This decrease in accuracy for profile poses are likely due to the small amount of texture information contained in profile images.

Another observation worth mentioning is related to the window size used to compute each descriptor or, equivalently, the amount of blur applied to different scales of the original image. It was observed that as the window size increases, there is an optimal value in the ultimate classification accuracy of siblings and non-siblings, as is illustrated in Figure 6.3. Empirically, the width of the window that optimizes the final accuracy was found to be approximately 40% of the distance between anchor points Δ (the exterior corners of the eyes for frontal images and nasion and pogonion for profiles – *cf.* Figure 3.4). This value was observed for both HQ_{faces} and LQ_{faces} and different values of Δ.

Another relevant observation emerged from the analysis of pruned SIFT features. After the feature selection process, we wanted to observe which landmarks were more often selected throughout different Pair Dataset (PDS). For this experiment, five PDSs were built using the positive siblings pairs of each Individual Dataset (IDS) and an equal number of randomly chosen negative pairs. Then, SIFT descriptors were extracted and were processed by the FS algorithm. Their final contributions to each of the five PDSs were summed and the result is shown in Figure 6.4. It is quite clear that the portions of the face contributing the most to the final sibling classification are the eyes and the mouth.

These observations led to the investigation of more robust features that, although still based on SIFT, considered all color channels, as explained in the next section.

6.1.4 Color descriptors

As described in Section 4.3.3, we tested different color descriptors. Experiments following the same protocol explained in the previous section yielded the results shown in Figure 6.5, where the squares highlight the regions where the color descriptors contribute the most to the final classification using five randomly chosen datasets..

Finally, we considered the features presented in Chapter 4 to obtain the most relevant

results that are presented in the following sections.

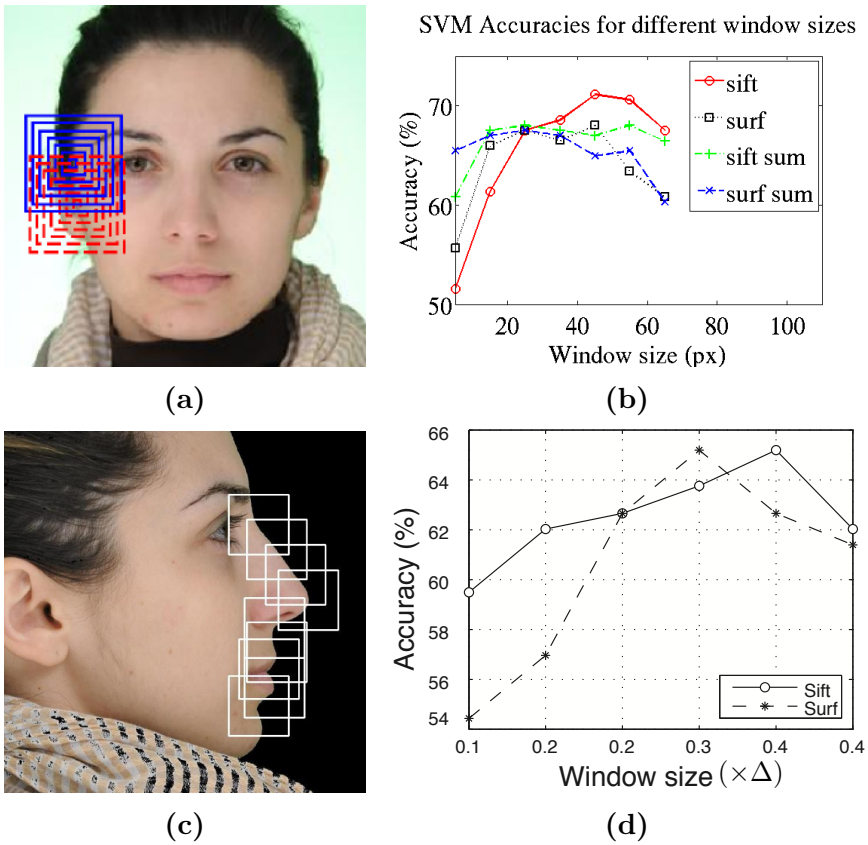


Figure 6.3: SVM classification for datasets composed of sift and surf descriptors computed using different window sizes. (a) Frontal expressionless image of one individual from *HQ-f*. (b) SVM accuracies for SIFT and SURF using different window sizes. (c) Expressionless profile image from *HQ-fp*. (d) SVM accuracies for SIFT and SURF computed with different window sizes.

Figure 6.4: Features selected after the two-step feature selection applied to the SIFT obtained from frontal HQ images.

6.2 Discriminating between siblings and non-siblings

The performances of automatically classifying pairs of siblings are presented in this section. By analyzing such close kins with high relatedness coefficient ($r = 0.5$, *cf.* Figure 2.1), the goal is to obtain a robust basis in the investigation of kinship recognition, given that there are only a few studies in this area. Further steps consist in gradually taking kins with lower relatedness into account. For instance, when adding parent-child pairs to the analysis, the age difference is very likely to increase the complexity of the problem.

Recalling from Chapter 3, the siblings classification investigation is performed using the following image databases:

- HQ_{faces}, containing high-quality images of siblings;
- LQ_{faces}, composed by low-quality images of siblings (mostly celebrities) gathered from the internet.

Therefore, for the sake of clarity, all relevant results obtained in simulations and compiled into this chapter are organized as follows. Firstly, Section 6.3 shows the human panel classification of siblings. The human's judgment in classifying siblings is fundamental to assess the performance of the developed classifiers. Secondly, Section 6.4 shows the performance of a commercial face identification software applied to siblings recognition (both HQ_{faces} and LQ_{faces} datasets). This analysis was a first attempt in analyzing if the kinship recognition can be tackled by a traditional identification algorithm. Finaly,

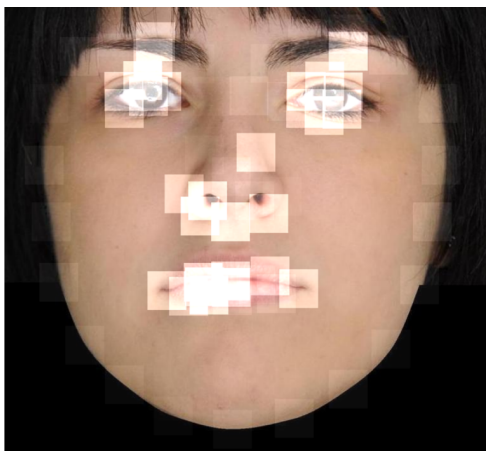


Figure 6.5: Features selected after the two-step feature selection applied to the Local Color Descriptor obtained from frontal HQ images.

a more robust classifier is developed using all features discussed previously in Chapter 4, and are presented in Section 6.5.

6.3 Human classification

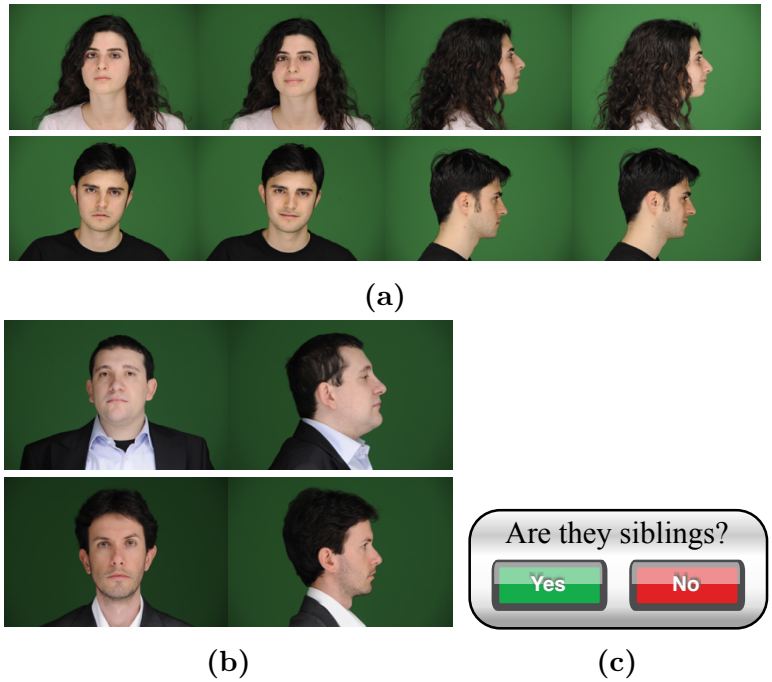
In order to asses the quality of our results, given the lack of touchstones, we intended to compare, on the same datasets, the capabilities of our automatic classifier with that of a panel of human raters. In the experiments, people were asked to evaluate if a couple of image sets depicted or not two siblings.

To this purpose, we presented the siblings pairs on an Internet site, to obtain Human Panel (HP) results. For each of the three Pair Datasets (PDSs), its pairs were presented one by one to each human rater in a random order, on a page where the images of each individual's Image Set (IS) were aligned in a row (*cf.* Figure 6.6). Members on the panel were informed that some of the pairs were siblings, but they were not told in what percentage. Members were asked to provide a YES (the two individuals are siblings) or NO (they are not) answer for each pair. In total, we collected 213.396 answers from 2.929 people; an average of 444 ± 0.5 answers for each pair was obtained.

Although the experiment was performed using an international website and many people outside Italy participated in it, the majority of voters were somehow associated to the Politecnico. This suggests that some siblings/non-siblings pairs could be correctly classified based on a previous knowledge of the voter about one or both individuals depicted in some questions. However, we did not expect to observe much bias in the ratings due to the large number of votes per sample.

The ratings showed a significant congruence between genders and age ranges of the volunteers. For instance, the Pearson's correlation between the classifications of the two genders was 0.88.

The results obtained in the HP experiment were examined based on a score $s \in [0, 1]$. For a given pair of subjects, s was computed from the fraction between the number of people guessing that the couple is related by the sum of all votes for that couple. And, if $s \geq 0.5$, humans classified the couple as two siblings, otherwise they guessed that the two individuals are unrelated. The percentage of correctly classified pairs are shown in Table 6.1 for each dataset belonging to HQ_{faces}. It was not possible to collect



(a)

(b) (c)

Figure 6.6: Example images presented in the questionaire answered by members of a human panel; (a) positive pair belonging to set *HQ-fps*; (b) negative pair belonging to set *HQ-fp*; and (c) question to be answered with YES/ NO options.

human expert ratings for the LQ_{faces} dataset, since it is composed mainly by well-known personages and hence likely to produce biased ratings.

The highest performance in Table 6.1 (75.22%) was achieved when the most information available was presented on the questionnaire, *i. e.*, when people were able to look at all possible poses and expressions of couple images (*cf.* Figure 6.6a). Based on this result, the automatic classifier should also perform better for this dataset specifically, and preferably, with accuracy higher than 75.22%. The following sections show that, indeed, it is possible to build an automatic classifier for siblings which outperforms the human ability in telling siblings from non-related couples.

Table 6.1: Classification results from the human panels experiment.

	HQ-f	*HQ-fp*	*HQ-fps*
Classification (%)	72.55	71.34	75.22
False Positive Rate (%)	0.98	1.22	1.77
False Negative Rate (%)	26.47	27.44	23.01

6.4 Commercial software

The concept of similarity of faces is much more encompassing than the concept of "identity". However, we believe to be interesting attempting to recognize pairs of siblings using an effective commercial identity recognition software, given the popularity of such applications. For this task, we selected the FaceVACS® Software Development Kit (SDK), supplied by Cognitec Systems [44]. A previous version of this software was tested in the Face Recognition Vendor Test (FRVT) 2006, obtaining excellent results in identity recognition [45].

When the SDK analyzes a pair of images, it provides a score value $s \in [0, 1]$. The higher the score, the higher the probability the images are from the same subject. Since siblings are likely to share facial attributes, one can suppose that the score between two siblings should be higher than the score between two unrelated people.

Indeed, FaceVACS can provide an initial insight when dealing with sibling images. We experimented FaceVACS on both databases, HQ_{faces} and LQ_{faces}, with all pairs of siblings and with an equal number of randomly selected non sibling pairs. As for the HQ_{faces} we applied the recognition software only to the expressionless frontal images of an individual, since FaceVACS claims to provide higher scores with such images. The histograms of the scores of pairs of siblings and non siblings show that no pair of non siblings scored higher than 0.4 for LQ_{faces} and 0.5 for HQ_{faces} (Figure 6.7). A fixed threshold might guarantee a null False Acceptance Ratio (FAR), but it strongly penalizes

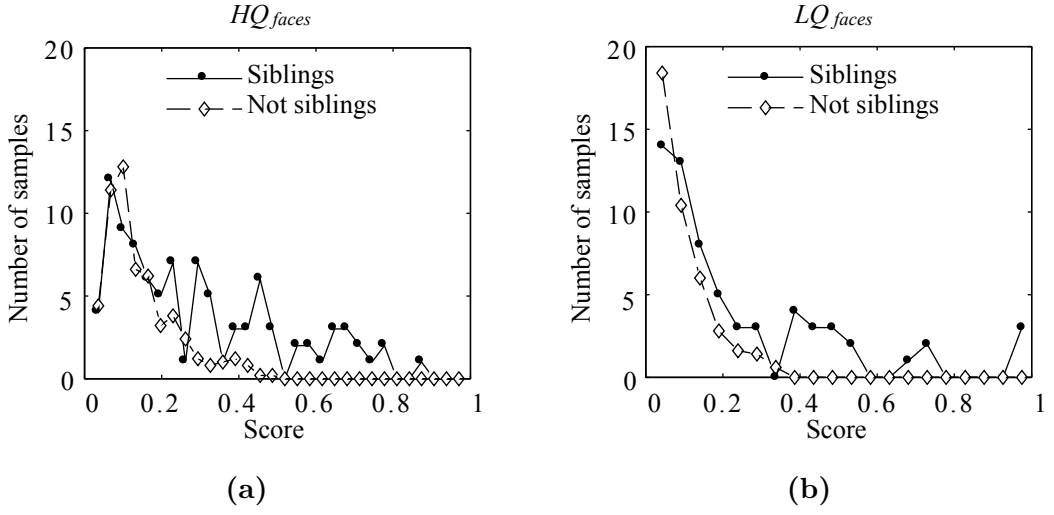


Figure 6.7: Histograms of scores provided by FaceVACs-SDK for pairs of siblings and non-siblings in: (a) HQ_{faces} and (b) LQ_{faces} (on the bottom).

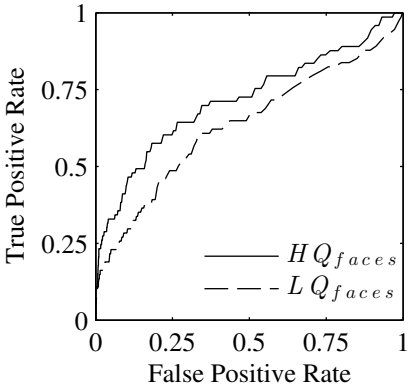


Figure 6.8: ROC curves obtained from applying FaceVACS to both LQ_{faces} and HQ_{faces}.

the False Rejection Ratio (FRR), since there are many siblings' pairs with scores lower than this threshold. For instance, to obtain a null FAR, the FRR is 78.12% for LQ_{faces}, and 82.47% for HQ_{faces}. The tradeoff between the false acceptance and the verification rate can also be verified with the ROC curves shown in Figure 6.8.

These results show that if the score of a couple of images is above these thresholds, they are likely to belong to siblings, otherwise, another algorithm should be used to make the decision. Concluding, the problems of identification and siblings' recognition appear rather different and an identity recognition software could not be able to provide effective sibling identification.

6.5 Automatic siblings classifier

As already explained in Chapter 3, the HQ_{faces} database was collected with the specific purpose of assessing the accuracy of the proposed approach on samples collected in controlled conditions. We recall that HQ_{faces} images were divided into three Individual Datasets (IDSs) according to their characteristics. In each IDS, given a set of N individuals and considering that there are no groups of three of more siblings, we have $N/2$ pairs of siblings and, potentially, $N(N-2)/2$ pairs of non-siblings. For each IDS, we created a Pair Dataset (PDS) containing all positive pairs, the $N/2$ pairs of siblings, and an equal number of randomly chosen negative, non-sibling, samples. For simplicity, in the following, the PDSs have the corresponding names of the IDSs used to build them (*i. e.*, *HQ-f*, *HQ-fp*, *HQ-fps*).

For each PDS the attributes described in Chapter 4 were extracted and used to construct different classifiers according to the method outlined in Chapter 5. Results were assessed using stratified five-fold Cross-Validation (CV) and, hence, the average classification rates of the SVM classifier over the different CV rounds are reported. For the SVM classifier, a Radial Basis Function (RBF) kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma||\mathbf{x}_i - \mathbf{x}_j||^2}, \quad \gamma > 0, \tag{6.1}$$

was used and parameters γ and C are optimized using a grid search defined by

$$\log_2 \gamma = -4, -3, \ldots, 1, \tag{6.2}$$

and

$$\log_2 C = -1, 0, \ldots, 3, \tag{6.3}$$

as suggested in [46].

In the first experiments, the classification accuracies were computed based on each attribute separately. Then, they were evaluated by characterizing each individual with three different groups of attributes. The first two, named GEOMETRIC and TEXTURES, gather together, respectively, all geometric and texture attributes, in order to understand if geometric or textural properties alone can discriminate properly between siblings and non-siblings. The last group, named ALL, concatenates all the described attributes to evaluate if the combination of characteristics of different nature provides a better solution to the problem. Concerning the FS process for these attribute groups, the set of feature variables used by the classifier for each PDS were obtained by first concatenating the optimal set computed applying FS to each attribute separately, and then again applying the FS algorithm to the resulting feature sets. This process can be seen as the selection of the "best of the best" features and, in preliminary tests, consistently provided better results than applying FS directly to the concatenation of all attributes.

The classification results are summarized in Table 6.2, where the classification accuracy (Acc. $\pm$ deviation) and the number of features selected by the FS algorithm (NFS) are reported, for each PDS and for each attribute or attribute group. These values are presented in Figures 6.9 and 6.10, to best illustrate the tendencies enumerated in the

Table 6.2: Classification results, showing for each PDS and for each attribute or attribute group, the classification accuracy (Acc) and the percentage of features selected by the FS algorithm (NFS). Higher accuracies are highlighted with brighter backgrounds.

	Pair Dataset (PDS)					
	HQ-f		*HQ-fp*		*HQ-fps*	
Attribute	Acc(%)	NFS (%)	Acc(%)	NFS (%)	Acc(%)	NFS (%)
1 NPOS	63.8 ± 6.3	34	69.6 ± 4.7	28	77.7 ± 5.6	22
2 SEGS	74.4 ± 4.7	62	77.2 ± 5.3	34	83.9 ± 4.5	48
3 ANGLES	71.9 ± 5.4	12	76.0 ± 7.1	28	85.7 ± 2.8	34
4 DTR	76.3 ± 9.6	18	79.7 ± 5.1	58	86.6 ± 7.2	24
5 PCA	73.8 ± 9.8	34	75.3 ± 11.2	20	76.8 ± 7.9	42
6 RIC-LBP	70.6 ± 5.3	72	67.1 ± 4.8	10	74.1 ± 6.6	16
7 RIGF	76.2 ± 4.8	44	74.1 ± 3.4	48	75.9 ± 5.4	20
8 LID	80.6 ± 6.7	20	76.6 ± 4.8	60	80.4 ± 2.3	34
GEOMETRIC	76.7 ± 7.4	25	84.2 ± 3.3	39	87.0 ± 4.4	38
TEXTURES	79.4 ± 6.7	13	79.7 ± 4.5	19	83.9 ± 5.0	29
ALL	81.9 ± 8.8	11	84.2 ± 5.3	17	88.4 ± 6.8	27

following remarks. The student's t test confirmed the accuracies mean values within a 5% significance level.

The following remarks can be made on the classification results summarized in Table 6.2:

1. The more information available, *i. e.*, the more poses analyzed, the higher the accuracy of the classifier. Profile and expressions significantly improved the classification results, as can be seen in Table 6.2, where results for set *HQ-fps* consistently outperform those for set *HQ-fp*, which are in turn better than those for set *HQ-f*. The only exception is represented by results on C-SIFT attribute, whose best performance was obtained on set *HQ-f*.

2. Frontal and profile images obtained on the average similar accuracies, exception made for textural descriptors and their combination, which performed bad. The weakness of the textural characterization of profiles is also underlined by the low number of selected features, which seems to indicate that just a very few regions around the facial landmarks contain relevant texture that might be useful for the proposed problem. Results showed that these points are the corners of the eye and mouth, and the nose wing contour.

3. Concerning geometric attributes, landmark positions alone did not provide signif-

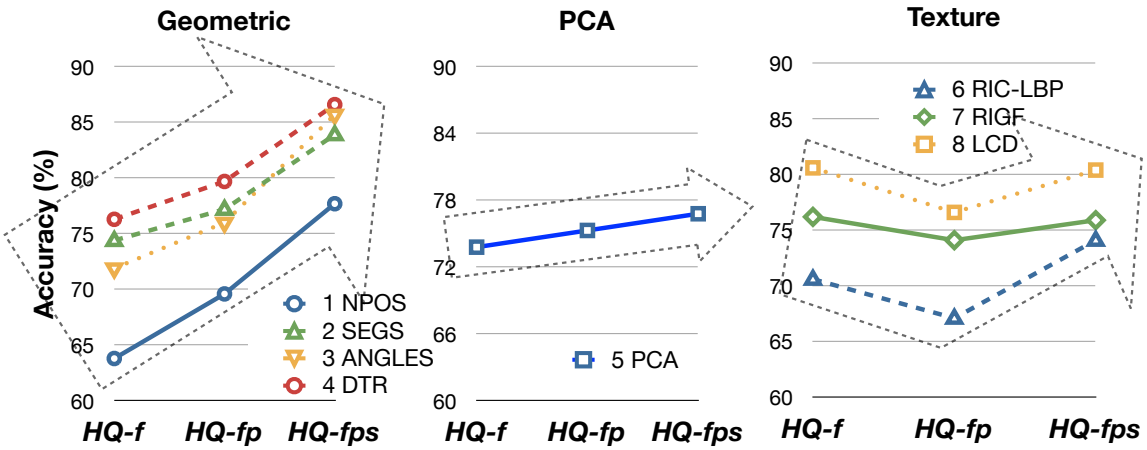


Figure 6.9: Illustration of the accuracy values of individual attributes from Table 6.2.

icant results. As expected, lengths and angles obtained similar results since they describe correlated information. Segment rations (DTR) were consistently better than any other geometric attribute, suggesting their superior capability in describing local similarities, and variations, of face shapes.

4. As for textural descriptors, C-SIFT obtained the best accuracies for all the PDSs, showing they are more suited to capturing the sibling characteristics. On the contrary, RIC-LBP provided, in most cases, poor performances.

5. The single attribute providing the best average result was DTR, immediately followed by C-SIFT; the holistic attribute (PCA) did not perform as well compared to them.

6. Grouped attributes performed better than their single components and the combination of all possible attributes provided the highest classification on all datasets. That is, the more heterogeneous the information, the better the performance. The only exception was TEXTURES applied to *HQ-f*, but the small difference with its best individual attribute (1.2%) can be explained in terms of the greedy nature of the heuristic FS algorithm.

7. The classifier achieved up to 88.4% of correct recognitions for set *HQ-fps*, 84.2% for set *HQ-fp* and 81.9% for set *HQ-f*. Although apparently good, it is not possible to directly evaluate their quality or to compare them with that of other approaches

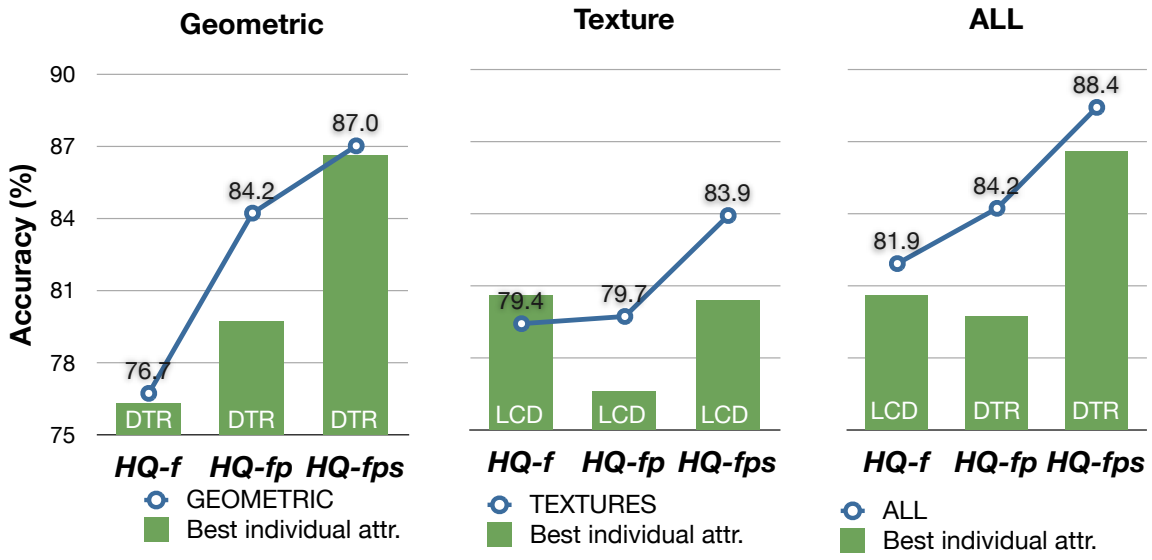


Figure 6.10: Illustration of the accuracy values of grouped attributes from Table 6.2.

(such as [16]) since they were obtained on different data. In the next subsection we propose such an evaluation through the comparison with the classification results obtained on the same data by a panel of human raters.

Analysis of the Feature Selection process.

As expected, FS always significantly increases accuracy. The general behaviour for different attributes is that the second FS step provides a major accuracy improvement over the first step (see as an example the results in Table 6.3). The large differences with the results computed on the initial feature sets, where no FS was applied and the classifiers were based on a much greater number of features than the number of observations, highlight as well the capabilities of FS to prevent overfitting.

Table 6.3: Classification accuracies (using ALL attributes group) applying SVM to (i) the initial feature set (NOFS), (ii) the reduced feature set extracted by mRMR (FS1) and (iii) the final feature set obtained by combining mRMR and SFS (FS2).

	HQ-f		*HQ-fp*		*HQ-fps*	
	Acc.	Diff.	Acc.	Diff.	Acc.	Diff.
NOFS	70.6	-	70.9	-	80.4	-
FS1	75.0	4.4	73.4	2.5	82.1	1.7
FS2	81.9	6.9	84.2	10.8	88.4	6.3

The FS process raises another question. Both mRMR and SFS are heuristics that are dataset dependent. This means that their application to different datasets are likely to produce optimal sets containing different features. How does this affect the generalization capabilities of the classifier?

Further experiments showed that the most significant features to discriminate siblings are relatively stable with respect to the pair dataset used. Furthermore, the features selected for a dataset, even if not optimal, still provide good classification accuracies on different datasets. In detail, we created, for each IDS, five more random PDSs, composed by an equal number of siblings and non-siblings. Then, for each PDS, we compared two classifiers built, respectively, (i) on the optimal feature set computed by applying FS to each PDS and (ii) on the (sub-optimal) feature sets computed from a different dataset, namely the corresponding PDS used to build Table 6.2. The average differences in their accuracies are low (few percentage points). Though further investigations with larger datasets would be necessary, these results are believed to support the intuition that the classifiers are quite robust with respect to the selected feature sets.

The analysis of the features surviving the FS pruning can provide some insights into the more relevant facial characteristics to recognize siblings. Grouping geometric attributes, NPOS variables are mostly discarded and DTR and ANGLES appear to be the most relevant. C-SIFT provides by far the most relevant contribution to the TEXTURE groups, while RIC-LBP can be considered marginal. In the ALL group, all types of features (*i. e.*, geometric, holistic and texture descriptors) are present, with a slight preference for C-SIFT and DTR and features from all image types are considered (*cf.* Figure 6.11). Although these results are dataset dependent, they suggest that (i) the combination of features with different characteristics provides a substantial improvement of the accuracy and (ii) information from different images of the subject is effectively combined to improve the classification results.

Influence of age difference

One interesting result is that the classification of pairs is completely independent from the age difference of their individuals. For example, its correlation with the classification variable in the ALL experiment falls within range [-0.050,-0.008] for different PDSs (*cf.* Table 6.4). However, the average age difference was rather low (4.6 years) and experiments

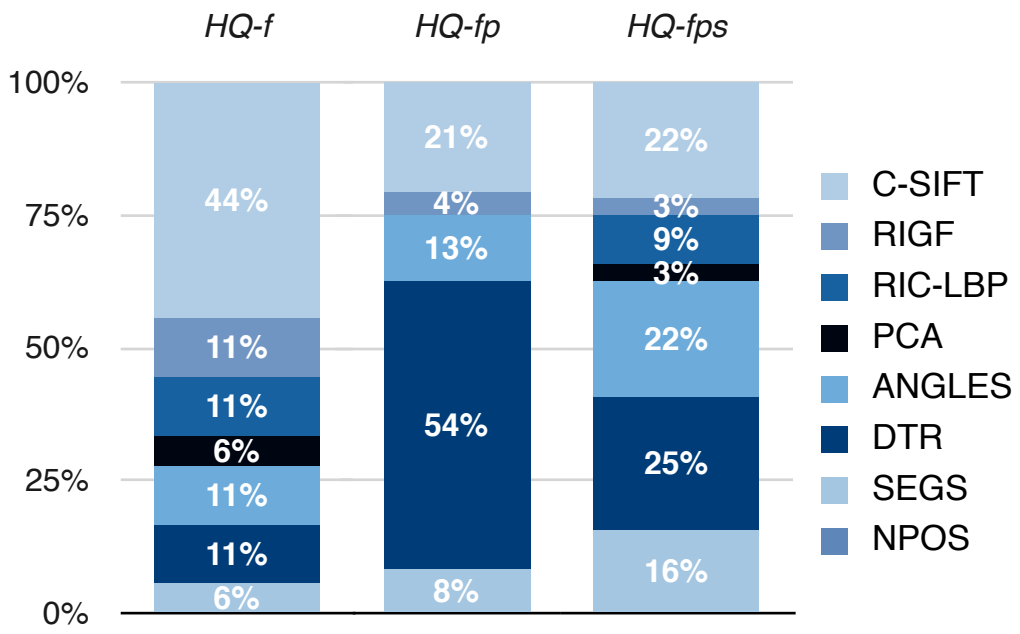


Figure 6.11: Proportions of chosen features selected from SFS applied to ALL attribute group.

on more challenging datasets are required to fully understand the influence of age on the final performances.

Other experiments

In order to understand if these results are due to the discriminating capabilities of chosen features or to the classification method, the SVM was compared to different classification techniques. Results have shown that SVMs performed consistently better (*cf.* Figure 6.12). But the average differences of few percentage points obtained on the three PDSs, with the ALL attribute group, by k-Nearest Neighbors (-6.0%), Bagging Trees (-5.5%), and Random Forests (-2.8%), suggest that is the design of the features that really makes the difference.

Finally, individual poses were tested separately (frontal, profile, expressionless and smiling faces) from *HQ-fps*, besides concatenating all of them. For each image type, we built a Pair Dataset Frontal (PDF) using the same negative and positive pairs of individuals used in the *HQ-fps* PDF. The best results obtained are very similar (between 80.4%

Table 6.4: Correlation between the classification variable and age differences.

Dataset	*HQ-f* (%)	*HQ-p* (%)	*HQ-fp* (%)	*HQ-fps* (%)
ALL	-0.020	-0.050	-0.008	-0.030

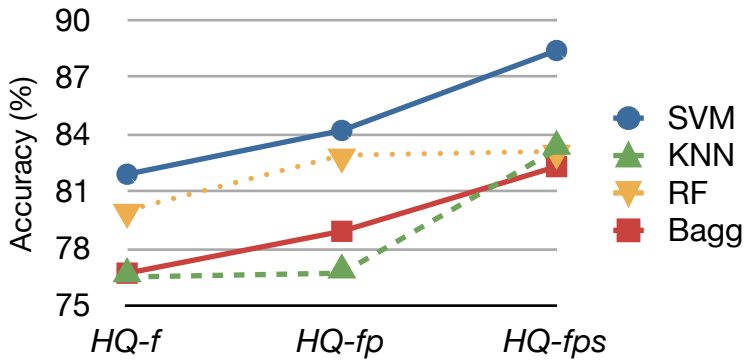


Figure 6.12: Results of other classification techniques.

and 84.8%), with expressionless images scoring slightly better than smiling faces (+4.4% and +2.0% for, respectively, frontal and profile faces). The comparison with 88.4% of the *HQ-fps* PDF shows that their combination is indeed effective to improve the classification accuracy. Five-fold cross-validation is performed in all cases and the parameters used were the default parameters for each classifier implemented by the MatlabArsenal package, whose specifications can be found in [47], except for SVM, where the parameters used were specified in Section 5.

Another comparison can be made with the results obtained by FaceVACS. The classification of *HQ-f* images, using a threshold of 0.5 on the FaceVACS score, had an accuracy of 55.5%, which gives a further indication that the commercial face recognition software cannot suitably tackle the sibling identification problem.

Comparing automatic and human classification

In order to assess the quality of these results, given the lack of touchstones, the performance of the automatic classifier was compared with that of a panel of human raters who were asked to evaluate if a pair of image sets depicted two siblings or not (*cf.* Section 6.3).

In order to perform a meaningful comparison with the classifier, we transformed the average ratings of the human panel (HP) into the value that obtained the majority of

Table 6.5: Classification accuracies per pose of faces belonging to set *HQ-fps*.

	f	*p*	*fs*	*ps*
GEOM.	79.5	84.0	80.4	82.1
TEXT.	82.1	73.2	78.6	75.9
ALL	84.8	80.4	77.7	79.5

votes for each pair. The best results obtained by the automatic classifier (based on the ALL attribute group) outperformed that of the HP on all PDSs. The differences in accuracies were +9.3% for *HQ-f* dataset and +12.9% and +13.2% for, respectively, *HQ-fp* and *HQ-fps* (Table 6.6).

These results are believed to emphasize the validity of the classification method.

Generalization properties of the classifier

Lastly, one last question must be answered: what are the generalization properties of the classifier? Or, in other words, how well does it behave on unseen data? Theoretically, it should behave well since several researches in literature support the claim that SVM has good generalization characteristics. These properties stem from the maximization of the margin of the hyperplane separating the two classes that, in turn, leads to the minimization of the generalization error. We tried to answer the initial question in the following way:

- we trained a classifier with all samples in the frontal pair dataset rated by the human panel (*HQ-f*); the classifier was based on the 16 variables from the ALL attribute providing the best classification accuracy for this dataset (Table 6.2);
- we built a test set from LQ_{faces}, containing the 98 siblings pairs and an equal number of randomly chosen non-siblings pairs;
- we classified the test set.

With such experimental settings, one point to discuss is about the color descriptors used in the classifier. As stated in Section 4.3.3, C-SIFT was the local descriptor providing best results for the HQ_{faces} datasets. However, C-SIFT is only invariant to light intensity

Table 6.6: Comparison of automatic and human classification of different datasets.

	Accuracy (%)		
	HQ-f	*HQ-fp*	*HQ-fps*
Human classification	72.6	71.3	75.2
Automatic classification	81.9	84.2	88.4
Difference	9.3	12.9	13.2

changes and, probably, it is well suited to classifying such images since they were shot in a controlled environment.

When dealing with images with more general properties, such as the ones in LQ_{faces} dataset, color descriptors with higher invariance are likely to provide better results. To verify this assumption, we built different classifiers for *HQ-f* again grouping all the attributes but substituting C-SIFT with another descriptor. In each case, we applied the "best of the best" technique to compute the optimal feature set.

The best classification accuracy was obtained with *rgb*-SIFT (78.5%) and it was slightly better than that obtained by *opponent*-SIFT (77.3%); it should be noted that *rgb*-SIFT is the only descriptor introduced in Section 4.3.3 that is invariant to all properties, while *opponent*-SIFT is invariant to the first three properties. These two results improved, as expected, the 75.9% obtained with C-SIFT (see Table 6.7). All those values are lower than the accuracy obtained from the classification of the *HQ-f* dataset (81.9%), but this difference can be explained in terms of the lower quality of the images in LQ_{faces}. Varying illumination during shooting, a not always exactly frontal position of the subject, the presence of expression in the images are all factors that heavily influence the results. Although it was not possible to collect human expert ratings for the LQ_{faces} dataset, since it is composed mainly of well-known personages and hence likely to produce biased ratings, the accuracies obtained can be compared to those obtained by the HP when classifying a dataset of frontal images (72.6%).

Table 6.7: Cross and within database classification accuracies on LQ_{faces} with ALL attribute group and different color descriptors.

Descriptor	Cross-DB acc. (%)	LQ_{faces} acc. (%)
C-SIFT	75.9	79.1
Opponent-SIFT	77.3	80.2
rgb-SIFT	78.5	81.2

A further verification of the generalization capabilities of the proposed approach, the method discussed here was compared to a previous investigation [16], which is the only work explicitly dealing with sibling classification. The method was applied to the VADANA dataset [48] following the same experimental protocol defined in [16]. Since authors were not able to provide us the exact list of pairs used in their tests, we analyzed a pair dataset that follows similar restrictions obtaining an accuracy of 78.9%, which can

be compared with the 75.6% of [16].

Part II

Classification of Parent-Child Pairs

Chapter 7

Parent-Child Classification

In the previous part of this thesis, the classification of siblings was discussed. This first investigation consisted in i) assembling two image databases; ii) normalizing all images; iii) extracting features from the single images; iv) combining the representative vectors of couples; and, finally v) classifying the datasets and selecting most descriptive features. This first research suggested the effectiveness of the classification scheme associated with the feature selection process. This methodology indicated which features are more prominent in the classification of kins and which allowed a broadening of the kinship recognition investigation by analyzing parent-child pairs. This second part deals with the automatic discrimination between parent-child pairs from unrelated individuals.

To this end, a database containing photographs of parent-child pairs was used. It was collected in a previous work in the literature, by Fang *et al.* [15], which also provided results from human panels on the same dataset. This was done to provide a reliable ground for comparison, *i. e.*, if the methodology presented in this work is able to outperform the results obtained in the previous work [15], then this is a strong suggestion of the effectiveness of the features extracted along with the classification approach.

Although very similar to the discrimination of siblings presented in Part I, the differentiation of parent-child pairs is done in a slightly different manner. This is because the experiments with siblings provided consistent insights regarding the method. Specifically, the siblings experiments were able to tell that some features are not satisfactorily descriptive to be used in the method. In addition, another textural descriptor was used, namely Weber Local Descriptor (WLD), which showed highly descriptive capabilities. Also, previous results have shown that classification techniques other than SVM provide

lower accuracies.

The following sections are organized as follows. Section 7.1 explains the work of Fang *et al.* [15] in details and its achievements. Section 7.2 explains the new feature used in the parent-child classification problem and its characteristics. The results are presented and discussed in Section 7.3 and conclusions of all work produced in this thesis are drawn in Section 9.

7.1 Parent-child database

Previously, two databases of facial images were described, namely HQ_{faces} and LQ_{faces}. Both were used in the automatic classification of siblings pairs as shown in Chapter 6.

A natural second step to deepen the analysis of kins is trying to classify parent-child pairs. To do that, the approach was finding an already collected database of parent-child pairs from previous works by other authors in the literature. Starting from an already prepared database provides two main advantages: (*i*) sparing time in collecting and labelling images; and (*ii*) being able to reliably compare the approach with different algorithms/ methods.

Therefore, previous works in the literature dealing with databases composed by parent-child image pairs were surveyed but, the method presented in this work could not (or should not) be applied to some of them, as explained in the following.

The parent-child database collected by Jiwen Lu *et al.* [49] is composed by faces

Figure 7.1: Database of parent-child pairs collected over the internet by Fang *et al.*.

cropped in a way that precludes the very first step of method presented here, the detection of facial landmarks (*cf.* Section 3.3). The one associated to the work of Xia *et al.* [50] contains many grayscale images, especially those of young parents, compromising the approach which relies in color descriptors (*cf.* Section 4.3.3). Therefore, the database collected by Fang *et al.* [15] was selected, since it is befit to our method, providing a reliable ground for comparison. In this section, their database and achievements in classifying the parent-child image pairs are discussed.

The database consists of 300 images collected from the Internet (150 parent-child pairs) and henceforth is referred to as PC_{faces} (for Parent-Child face pairs). The photographs were taken from several public celebrities of different age, gender and race in slight different poses (mostly frontal), illumination conditions and expression (some neutral but often smiling). Examples of images in this DB are shown in Figure 7.1. The database is available online[1] and contains 50% Caucasians, 40% Asians, 7% African Americans, and 3% others; 40% of the 150 images are father-son pairs, 22% are father-daughter, 13% are mother-son, and 26% are mother-daughter. They used a simplified Pictorial Structure Model (PSM) to obtain the positions of four facial landmarks on each image (right and left eyes, nose and mouth), which served as a base for extracting features from the images: (*i*) color, central or most commonly occurring color on landmarks; (*ii*) facial parts, sub-windows centered on the keypoints; (*iii*) facial distances, euclidean distances between landmarks normalized by global face dimensions; and finaly (*iv*) gradient histograms. Ultimately, each image is represented by a vector containing 22 different features associated to different measurements yielding a final vector of dimension much higher than 22. They used Sequential Forward Selection (SFS) [40] to derive a set composed by features providing best accuracy. This approach consists in first taking one feature that solely provides the highest accuracy. Then, another feature is added such that, when considered in conjunction with the previously selected, increases the accuracy the most. This process continues till a maximum accuracy is achieved. Since this technique relies on accuracy performance, it depends on the classification method chosen. The classification was performed using KNN with k=11 and SVM with a Radial Basis Function RBF kernel. The separation between training and test sets was implemented using 5-fold cross-validation. Finally, by applying SFS, a 10 dimensional vector composed

[1]Although the authors reported the use of 150 pairs, there are 144 pairs available online, *i. e.*, 288 pairs.

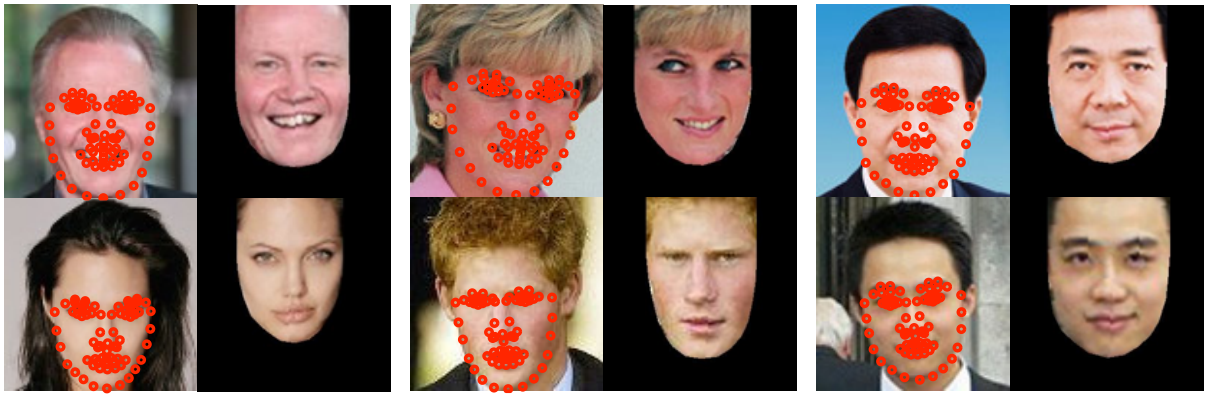


Figure 7.2: Examples of three parent-child pairs belonging to PC_{faces}. Images on the left show originals with annotated landmarks, whereas images on the right show the normalized faces.

by six-features is constructed for representing a couple. Overall classification accuracies of 70.67% and 68.60% were achieved with KNN and SVM, respectively. These results outdid the classification performed by a panel of human raters, which achieved 67.19% of correct classification of parent-child pairs.

7.2 Weber Local Descriptor (WLD)

Chen et al. [51] developed an image descriptor coined WLD based on Weber's law, which states that the change in a stimulus that will be just noticeable is a constant ratio of the original stimulus. This relationship can be expressed as

$$\frac{\Delta\sigma}{\sigma} = k, \tag{7.1}$$

where $\Delta\sigma$ represents the increment threshold (just noticeable difference for discrimination); σ represents the initial stimulus intensity and k signifies that the proportion on the left side of the equation remains constant despite variations in σ term.

The *Weber fraction* given by Eq. (7.1) can be intuitively understood by considering that a given variation in illumination intensity is more perceivable by the human eye in a dark room than in a bright environment. Taking this into consideration, the WLD is based on two components; *1*) differential excitation; and *2*) gradient orientation, whose general ideas are explained in the following.

Considering an $M \times N$ grayscale image I and a 3×3 region of pixel levels p, centered

on pixel p_n

$$\begin{bmatrix} p_1 & p_2 & p_3 \\ p_4 & p_n & p_5 \\ p_6 & p_7 & p_8 \end{bmatrix}, \tag{7.2}$$

the differential excitation ε_n is given by

$$\varepsilon_n = \tan^{-1}\left(\frac{1}{p_n}\sum_{i=1}^{8} p_i - p_n\right) \tag{7.3}$$

and the gradient orientation γ_n by

$$\gamma_n = \tan^{-1}\left(\frac{p_4 - p_5}{p_7 - p_2}\right), \tag{7.4}$$

for all $n = 1, \ldots, MN$ pixels of image I.

The first component aims at representing the Weber's law by mean of pixel values, i e., by dividing the summation of the differences between neighbors and central intensity by the central pixel value. The numerator in (7.3), containing intensity differences, represents the stimulus variation, whilst the denominator (central pixel value) represents the original stimulus. The differential excitation varies within the interval of the arctangent function, *i. e.*, $\varepsilon_n \in [-\pi/2, \pi/2]$. Image regions with high frequency variations have corresponding ε_n near $-\pi/2$ and $\pi/2$, whereas values around zero correspond to low frequency variations.

The second component, or gradient orientation, also lies in range $[-\pi/2, \pi/2]$. However, these values are shifted to range $[0, 2\pi]$ by analyzing the signs of the numerator and denominator of the arctangent's argument in Eq. (7.4). This is done to avoid the lost of directionality caused by the arctangent function. In this way, the gradient orientation represents the directionality of intensity variation.

Two examples of computation of WLD components are shown in Figure 7.3. Both components are obtained from the original images shown in Figures 7.3a and 7.3d. The differential excitations are shown in Figures 7.3b and 7.3e; the gradient orientations components are shown in Figures 7.3c and 7.3f respectively, where corresponding ranges are scaled to $[0, 255]$ for the purpose of visualization.

The differential excitation employs the advantages of LBP in computation efficiency and smaller support regions, whilst gradient orientation employs the advantages of SIFT

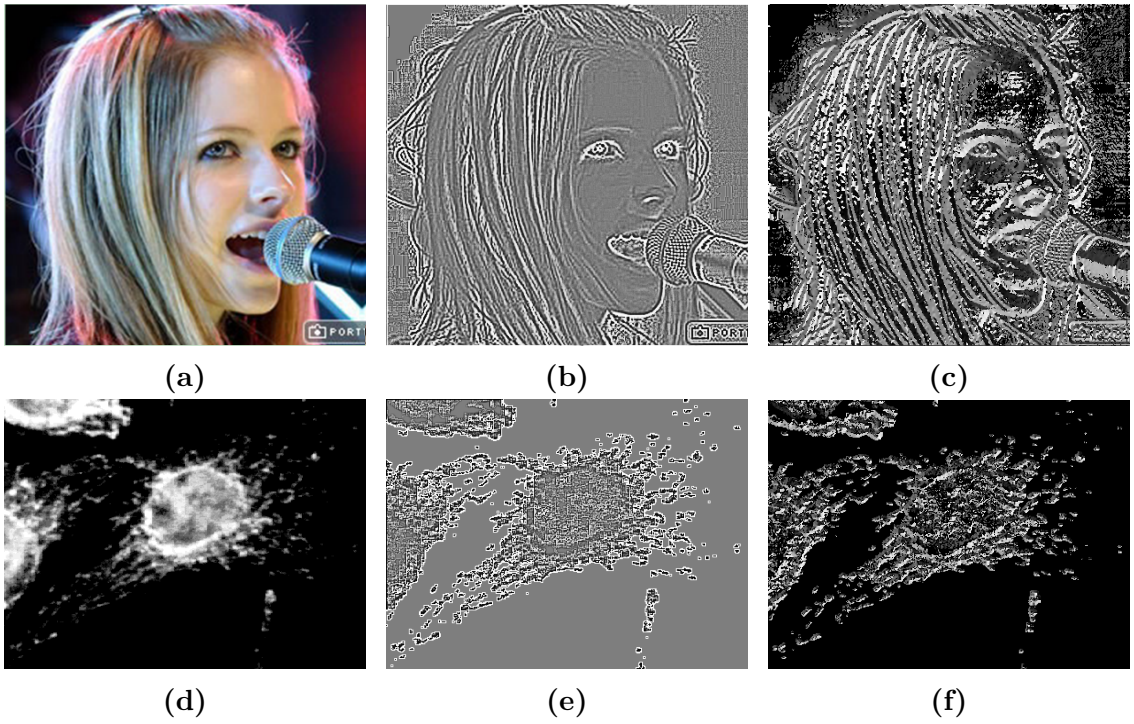


Figure 7.3: Two examples of the computation of the Weber Local Descriptor components: (a) and (d): original images; (b) and (e): differential excitations; (c) and (f): gradient orientations. Images (b), (c), (e) and (f) are scaled to $[0, 255]$ **for visualization purpose.**

by using the gradient orientation. Since both components are computed for each pixel in an image, WLD is a dense descriptor, differently from SIFT, which detects candidate keypoints prior to actually computing descriptors. Moreover, WLD do not depend on image size, although multi-scale analysis can be performed by varying the size of window (7.2).

Quantization and the WLD vector

The differential excitations and gradient orientations computed on each pixel of the $M \times N$ image I using Equations (7.3) and (7.4), are quantized to generate the final descriptor vector $\mathbf{h}$. This quantization is illustrated in Figure 7.4, where the possible excitation and orientation values are quantized into L_ε and L_γ levels, respectively.

Gradient orientation values γ_n that are greater or equal than a lower boundary $\lambda_{j,l}$ and less than an upper value $\lambda_{j,u}$ are quantized into the level represented by colum j. Similarly, the differential excitation values ε_n that satisfy $\lambda_{i,l} \leq \varepsilon_n < \lambda_{i,u}$ are quantized into row level i. Intuitively, each cell $h_{i,j}$ contains the number of pixels that have differential excitation and gradient orientation values inside ranges represented by i and j, respectively. More formaly, each cell value is given by

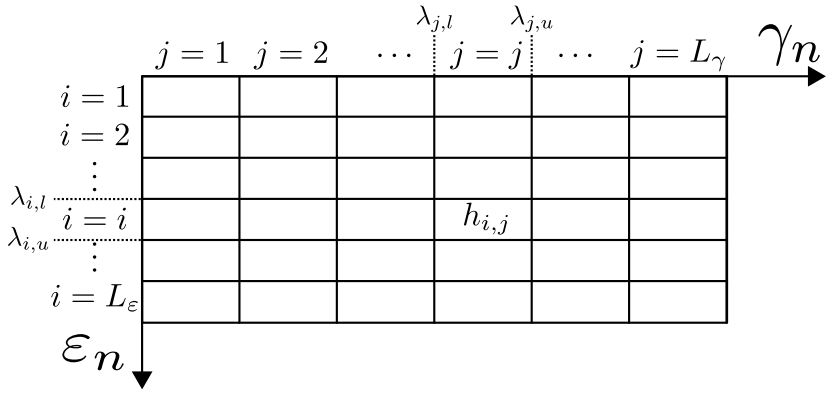


Figure 7.4: Weber Local Descriptor matrix.

$$h_{i,j} = \sum_{n=1}^{MN} f_{i,j}(\varepsilon_n, \gamma_n), \tag{7.5}$$

where

$$f = \begin{cases} 1, \text{ if } \lambda_{i,l} \leq \varepsilon_n < \lambda_{i,u} \text{ and } \lambda_{j,l} \leq \gamma_n < \lambda_{j,u} \\ 0, \text{ otherwise} \end{cases} . \tag{7.6}$$

The final feature vector $\mathbf{f}_{WLD}$ is then obtained by a concatenation of the $h_{i,j}$ values.

In this work the WLD is computed using the same parameters as the authors, *i. e.*, by quantizing the gradient orientations in 8 bins and the differential excitations in 120 bins, allowing a feature vector $\mathbf{f}_{wld} \in \Re^{2880}$ from each image or image region. An WLD feature vector example, computed from the original image show in Figure 7.3a can be seen in Figure 7.5.

Experimental results presented in [51] have shown that WLD outperforms LBP and Gabor descriptors based on its performance on classifying different benchmarks.

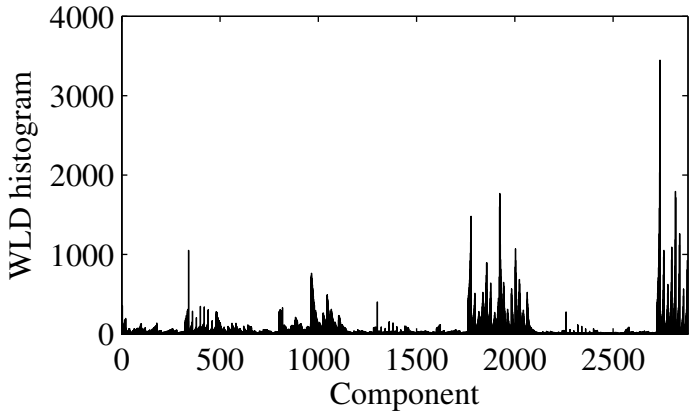


Figure 7.5: The final Weber descriptor vector $\mathbf{h} \in \Re^{2880}$ (or histogram) computed from the image shown in Figure 7.3a.

7.3 Results of parent-child classification

As one step further in the development of a multi-class discrimination algorithm, pairs of parent-child individuals were analyzed. Recalling from Chapter 7.1, Fang *et al.* [15] collected from the Internet various images of parents along with their corresponding children to perform kinship analysis. They obtained a classification accuracy of 70.67% in the classification of parent-child pairs, which outdid the 67.19% performance of human raters. In order to provide further insight in the effectiveness of the algorithm discussed in this thesis, it was also applied to the parent-child dataset collected in [15].

Here, as before, the accuracies based on the classification of each separate attribute was first analyzed. Attributes providing low accuracy were discarded. Then, they were evaluated by characterizing each facial image with three different groups of attributes: GEOMETRIC, grouping the geometric attributes, TEXTURE, combining textural information, and ALL, concatenating all the described attributes. The results are summarized in Table 7.1 and organized by attribute, or attribute group, and by classification algorithm (SVM vs. Random Decision Forest (RDF)). Again, to improve the understanding of the accuracies, the same values are illustrated in Figure 7.6. Other classifiers like KNN where not tested, since they did not perform well in the siblings classification problem, as can be seen in Figure 6.12. Results were assessed using stratified five-fold cross-validation (CV), and, hence, the average classification rates of each classifier over the different CV rounds are reported.

The following remarks can be drawn:

- concerning single attributes, textural features have an higher discriminative power than geometric ones, with WLD obtaining the best performances (78.0% with

Table 7.1: Accuracy results. For each attribute and each classification algorithm, we show the percentage of correct classifications and, in brackets, the optimal number of variables selected by the FS process.

	SVM (%)	NFS (%)	RDF	NFS
SEGS	68.2	36	60.1	40
RATIOS	73.1	26	59.3	175
ANGLES	68.9	60	57.2	100
C-SIFT	74.1	28	66.3	62
WLD	78.0	56	70.6	250
GEOMETRIC	74.3	13	65.4	175
TEXTURE	80.1	44	76.1	150
ALL	81.8	28	77.5	150

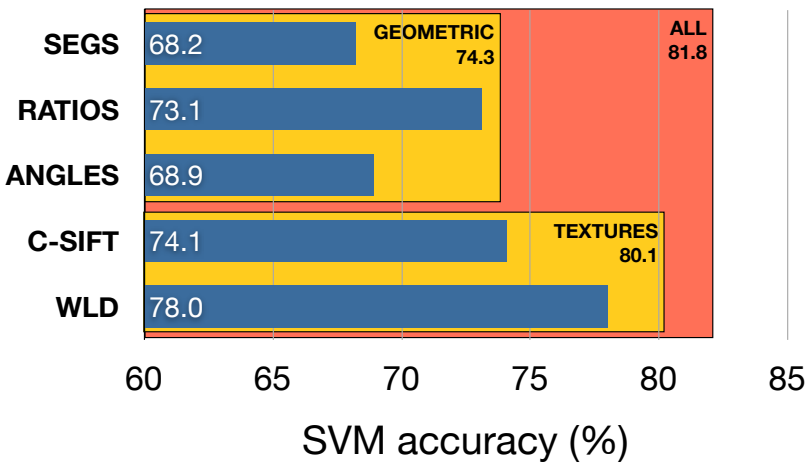


Figure 7.6: Classification accuracies of individual and combined features for parent-child pairs.

SVM);

- the more heterogeneous the information, the better the accuracies. As a matter of facts, grouped attributes performed consistently better than their single components, and the best accuracies were obtained for both algorithms considering all attributes together achieving 81.8% and 77.5% for, respectively, SVM and RDF;

- as for the classification techniques, SVM, in combination with a proper selection of the most relevant features, provides, in this specific problem, consistently better performances than RDF.

One expected result, not shown in Table 7.1, is that Feature Selection (FS) provides, in both cases a significant classification improvement (between 6% and 14%, for SVM, and 1% and 12% for RDF). As for the selection process, it is also interesting to analyze the distribution of features surviving the FS pruning for composite attributes. Figure 7.7 illustrates, for all attribute groups, the percentages of features types composing the optimal feature vector for both SVM (Figure 7.7a) and RDF (Figure 7.7b). In general;

- features from all attributes are chosen to compose the final vector, exception made for ALL group with SVM where ANGLES are discarded (which could be expected since they convey an information similar to SEGS);

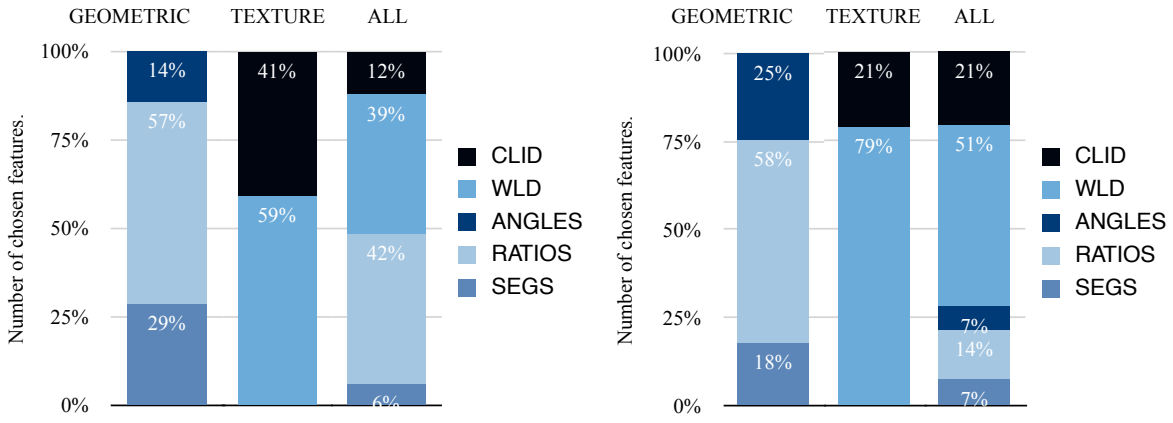
- RATIOS is the most relevant geometric attributes, suggesting its good descriptive capability;

- WLD features are found more relevant than CLID ones, but the latters also contribute reasonably when geometric features are added;

- texture features are preferably selected to compose the final dataset, especially with RDF.

Finally, these accuracies can be compared to those obtained by Fang *et al.* [15], who analyzed the same dataset with a different technique. The performance of their approach (70.69%) and that of a panel of human raters on the same data (67.19%) are improved by that obtained in our work with several single attributes, and outperformed by our best result (81.8%), achieved with the combination of all attributes.

In conclusion, the experimental results show that the combination of geometric and textural features, together with a proper selection of the feature variables, is indeed a valuable solution to the automatic Kinship Verification (KV) process, obtaining high classification accuracies and outperforming previous approaches on the same data.



(a) SVM: feature selection. **(b) RDF: feature selection.**

Figure 7.7: Feature selection applied to (a) SVM and (b) RDF: distribution of feature variables per type for different attribute groups.

Chapter 8

Siblings Classification Revisited

As stated in the Introduction (*cf.* Chapter 1), this work was organized into two parts, the first dealing with siblings and the second with parent-child pairs. The second part, assessing the problem of parent-child classification, considered the knowledge acquired in the siblings problem. Specifically, features providing low accuracies and low capacity to survive the Feature Selection (FS) pruning in the siblings analysis were discarded in the parent-child classification (NPOS, PCA, RIC-LBP and RIGF – *cf.* Chapter 7). Also, another textural feature was added to the parent-child problem, namely, Weber Local Descriptor (WLD), which provided good results.

Taking this into consideration, the analysis of siblings was revisited considering also the WLD feature, to analyze if the accuracies achieved in Part I could be improved. Moreover, the same analysis of the features surviving the FS pruning was performed in order to provide more insight on the features that are relevant for the classification of siblings.

Therefore, in this chapter are presented the results of the same methodology presented previously is revisited, *i. e.*, features are extracted and used to build the Pair Datasets (PDSs), which are then fed into the two-step Feature Selection (FS) process associated with the Support Vector Machines (SVM) classifier, as explained in Chapter 5.

8.1 New set of features

The knowledge acquired during the development of the siblings and parent-child analysis led to a selection of the attributes that, based on the accuracies results, are likely relevant

to Kinship Verification (KV).

Based on the results presented on past chapters, the attributes enumerated in Table 8.1 were considered. Also, the dimensions of each attributes are presented. The specification number of each attribute was maintained to avoid misunderstanding with the attributes used in the siblings problem, enumerated on Table 4.2. Notice that there are only geometric and textural feature groups.

8.2 Classification results

Following the same procedure presented before, all attributes were first tested individually, and then they were combined, composing a set of the "best-of-the-best" features, as explained in Section 5.2. Classification accuracies of individual attributes are presented on Table 8.2, where higher accuracies are highlighted with brighter background colors. Same as before, the percentage of features selected are shown for each Pair Dataset (PDS). The combined attributes were also tested in three different groups; (1) GEOMETRIC, concatenating all geometric attributes (SEGS, ANGLES and DTR); (2) TEXTURES, combining LID and WLD, and finally; (3) ALL, concatenating all features together. These combination results are also shown in Table 8.2. As before, one might notice that, in general, the more poses analyzed, the higher the performance. And, the more heterogeneous the data, *i. e.*, the combination of features of different nature yields, in general, better results. This can be best observed in Figure 8.1, which shows the classification accuracies for individual and combined attributes for Pair Datasets (PDSs) *HQ-f*, *HQ-fp* and *HQ-fps*. It is important to notice the following:

Table 8.1: Attributes dimensions.

	Attribute size (size of a feature × number of features)		Number of feature variables related to the attribute for an individual		
Attribute	Frontal (f) and frontal smiling (fs) images	Profile (p) and profile smiling (ps) images	*HQ-f*	*HQ-fp*	*HQ-fps*
2. SEGS	1×184	1×25	184	209	418
3. ANGLES	1×342	1×42	342	384	768
4. DTR	1×862	1×92	862	954	1908
8. LID	384×76	384×12	76	88	176
9. WLD	2880×76	2880×12	76	88	176

Table 8.2: Classification results, showing for each PDS and for each attribute or attribute group, the classification accuracy (Acc) and the percentage of features selected by the FS algorithm (NFS).

Attribute	*HQ-f*		*HQ-fp*		*HQ-fps*	
	Acc.(%)	NFS (%)	Acc.(%)	NFS (%)	Acc.(%)	NFS (%)
2 SEGS	74.4 ± 4.7	62	77.2 ± 5.3	34	83.9 ± 4.5	48
3 ANGLES	71.9 ± 5.4	12	76 ± 7.1	28	85.7 ± 2.8	34
4 DTR	76.3 ± 9.6	18	79.7 ± 5.1	58	86.6 ± 7.2	24
8 LID	80.6 ± 6.7	20	76.6 ± 4.8	60	80.4 ± 2.3	34
9 WLD	75.0 ± 4.5	34	75.3 ± 3.5	34	79.5 ± 6.6	18
GEOMETRIC	75.9 ± 6.4	33	82.7 ± 4.7	25	88.4 ± 2.5	57
TEXTURES	83.1 ± 3.8	89	83.5 ± 2.7	36	85.7 ± 4.3	42
ALL	83.1 ± 10.5	25	84.2 ± 4.8	16	92.0 ± 2.7	44

- The geometric features combined provided the same results as before, indicating that, indeed, the NPOS feature can be discarded with no considerable loss in performance.
- Textures provided consistently better results in comparison with the previous set of attributes analyzed in Table 6.2, showing that;
  - RIC-LBP and RIGF are likely to deprecate the performance.
  - The Weber Local Descriptor (WLD) performs well individually, contributing to a high performance when combined with attributes of different nature.
- The final classifier achieved $92 \pm 2.7\%$ accuracy of correct classification of siblings using ALL attribute group on *HQ-fps* Pair Dataset (PDS).

8.3 Selection of best features

Once again, an analysis of the features surviving the FS pruning might allow an insight of the features that are more likely to provide higher discriminative capabilities to the siblings identification problem. This analysis is performed by observing which are the features surviving the FS pruning when the algorithm is applied to the ALL attribute

group onto the *HQ-fps* Pair Dataset (PDS). It is important to recall that the ALL attribute group contains all features specified in Table 8.1, and the *HQ-fps* PDS contains images in frontal, profile, frontal smilling and profile smilling poses.

Figure 8.2 shows the proportions of features selected when the two-step Feature Selection (FS) process is applied to the ALL attribute group computed from the *HQ-fps* dataset. The following behaviors can be noticed:

- The features more resistant to the FS pruning are WLD, LID and DTR.
- There is a reasonably balanced preference for geometric and textural features for all PDSs. But, for *HQ-fps*, the geometric features dominate slightly composing 63% of the final dataset.
- There is a balanced choice between WLD and LID for textural features. For geometric features, on the other hand, there seems to be a preference for DTR, suggesting the descriptive capabilities of Delaunay Triangulation ratios between facial segments.

A comparison between the previous classification accuracies of siblings presented in Section 6.5 and the current classification of siblings can be appreciated in Figure 8.3.

8.4 Correct and incorrect classifications

In the investigation of facial analysis, it is pertinent to observe a few images correctly and incorrectly classified by the software. Indeed, by obtaining an insight about the correctly and missclassified samples, one might subjectively analyze if the algorithm is able to effectively represent facial similarity effectively or not.

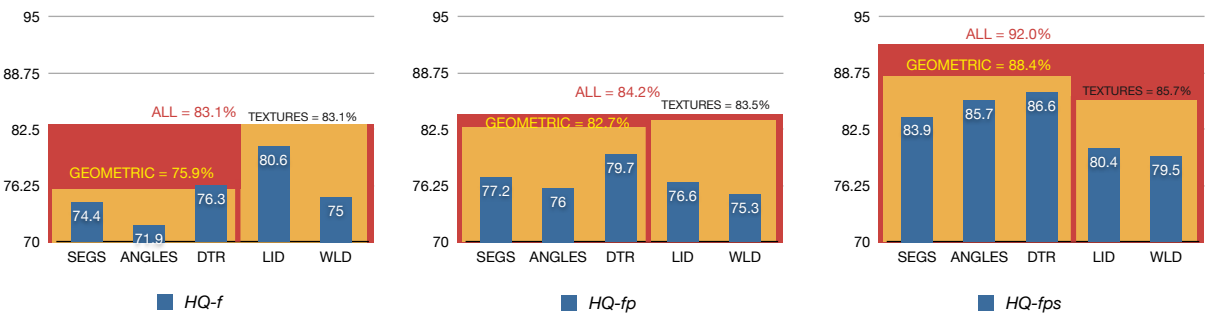


Figure 8.1: Classification accuracies for individual and grouped attributes for PDSs *HQ-f*, *HQ-fp*, and *HQ-fps* at left, center and right, respectively.

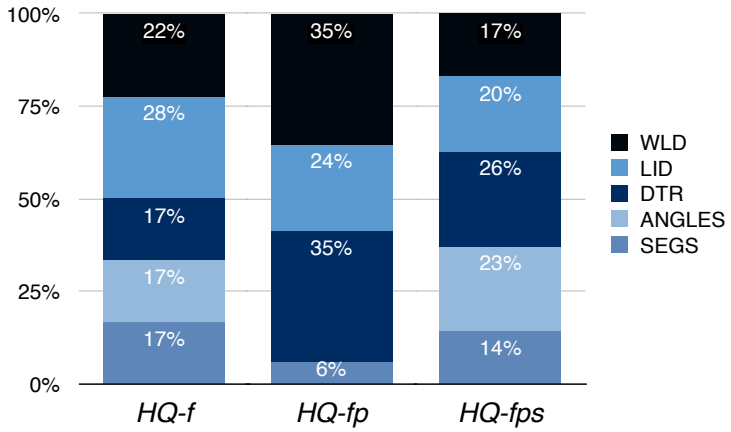


Figure 8.2: Proportions of features selected from ALL attribute group computed from *HQ-fps*.

To this purpose, consider Figure 8.4, which shows 5 samples from *HQ-f*. The first three samples are positive (siblings) and the last two are negative (non-siblings). In fact, observe that the two siblings in the first sample are identical twins. The SVM score columns presents the scores $s \in [0, 1]$ of the automatic classifier. The sample is classified as positive (siblings) if $s \geq 0.5$ and negative otherwise. Although the SVM scores are typically within range $[-1, 1]$ with the decision threshold equals zero, these scores where translated to facilitate comparison with the Human Panel (HP) scores. These are analogous to the SVM scores, *i. e.*, the decision threshold is 0.5. The closest both scores are to 0 and 1, the more confident is the classification.

The following observations can be drawn:

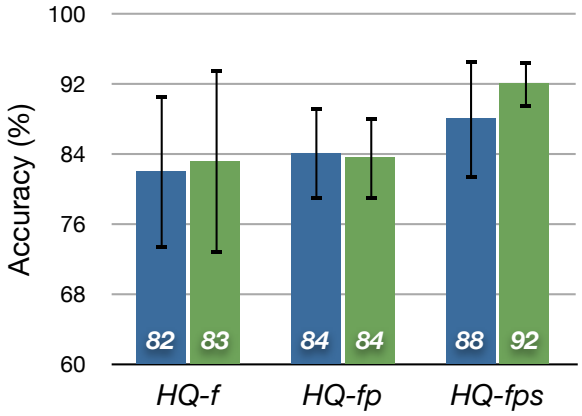


Figure 8.3: Comparison between previous and current accuracies in siblings classification.

Figure 8.4: Examples of correctly and erroneously classified samples.

1. Sample number 1 (positive). Both SVM and humans classified correctly, but SVM provided much higher confidence. This might be and indication that the software can perform well if applied to traditional identification.

2. Sample number 2 (positive). Only SVM classified this siblings' pair correctly, with high confidence. Notice the similarities in the faces of the individuals. Their eyes, nose, mouth and chin present similar shapes. Their eyes have the same color.

3. Sample number 3 (positive). SVM provided very poor (incorrect) classification in

this case. Humans correctly classified it with poor confidence.

4. Sample number 4 (negative). Both classified the sample correctly with reasonable confidence.

5. Sample number 5 (negative). SVM provided incorrect classification whereas humans correctly classified the sample with high confidence.

No negative pair was correctly classified by SVM and incorrectly classified by humans. This occurred because members participating in the HP experiments preferred to say that a sample was negative if they had "any" doubt about the brotherhood of the sample. This can be observed in the low false positive ratios presented in Table 6.1.

8.4.1 Processing times

It is relevant to briefly discuss the processing times involved in each step of the algorithm discussed in this work. They are enumerated in the following.

1. **Detection of facial landmarks and image normalization**. The time necessary to detect the 76 and 12 points located onto frontal and profile images is of the order of a few seconds, depending on the image size and the effectiveness of the Viola-Jones face detection [52]. Nevertheless, this processing time can be considered negligible when compared to the time of the features selection processing. The image normalization also has a negligible processing time when the whole algorithm is considered.

2. **Extraction of features.** From all features discussed previously, the more expensive to compute are the SIFT-based color descriptors, which took approximately 6-8 seconds to compute. The remaining textural features take a few seconds to process. The WLD for instance, needs 1.8 ± 0.5 seconds to compute.

3. **Classifier training**. This is, by far, the most expensive step in the whole algorithm:

    - The first feature selection step, mRMR, however, is fast since the samples are categorized using the mean $\pm$ deviation scheme. Besides, its implementation is available in Matlab® executable file (.mex), which performs better than

script files. It also needs only a few seconds to rank the 50 features with least redundancy between each other and most correlated to the classification variable.

- The second FS step, Sequential Forward Selection (SFS), however, is the most expensive one due to its iterative nature. Depending on the size of the feature set, it can last from 5 to 20 minutes long. During the development of this thesis, simulations would eventually last for days, due to iterative evaluation of different parameters, such as, for instance, the number of features ranked by mRMR.

4. **Classifier testing**. Considering that the SVM classifier is already trained and the final set of features was selected in the training phase, the testing phase is quickly executed. It consists in normalizing the two test images, extracting the features (which is performed in $\approx$ 10 seconds), mapping those features to a higher dimensional space using the optimized kernel and finding its distance to the separating hyperplane. This classification is quickly performed (few seconds) by the `libSVM` package [46], which is also available as a Matlab executable file (.mex).

Chapter 9

Conclusion and Future Work

This thesis has presented the results of a new research aimed at recognizing siblings' pairs from 2D color images with pattern analysis and image processing techniques. To this purpose, a database of high quality images of pairs of siblings shot in constrained poses and in a controlled environment was constructed, which will be made available to the community for further investigation on the subject. The ability of human observers to discriminate pairs of siblings and not siblings from images of this database was experimentally determined as well. Then, different facial attributes were extracted from available images, which can be divided into three main groups: (i) geometric information, related to the shape of the face or of its relevant features; (ii) holistic data, combining geometric and textural information; and (iii) local image descriptors, describing the image characteristics in the neighborhood of some salient facial points. Finally, different attributes and their groupings were used to build different classifiers, based on the integration of SVMs with a two-step Feature Selection process.

Results show that the combination of features of different natures is effective in achieving higher accuracies than those obtained by a panel of human raters, which were used as a basis for comparison in order to assess the quality the method. The generalization capabilities of the proposed approach are shown by the satisfactory accuracy of the classification of a dataset of heterogeneous images collected from the Internet. All relevant results presented in this thesis were published in [5, 6, 7]. The first observation that *eigenfaces* could be effective in discriminating siblings was presented in [5]. The relevant results obtained by analyzing all features described in Part I were presented in [6], whereas the results in discriminating parent-child pairs (*cf.* Part II) were shown in [7].

9.1 Future work

Hopefully this research and the collected database of high quality facial images will constitute a basis for future studies in the field of automatic kinship recognition. Several further steps are possible:

- A larger siblings database would be important for verifying the algorithms proposed and strengthening the results. For instance, it will be relevant to collect siblings of different ethnicities. As stated in Chapter 3, the database collected contains only images of Caucasians, which is unbalanced to deal with real-world databases, since individuals present a large diversity in skin/hair/eyes colors and geometry of facial parts, both locally and globally.

- It will be interesting to investigate the relevance to kin identification of very strong similarities limited to restricted face areas (like eyes, nose, mouth), which is considered a kinship clue in everyday life. This is a sensitive approach, since kins can have strong or weak similarities of specific facial parts. For instance, the eyes of a daughter can be very similar to her mother's eyes but, on the other hand, very dissimilar to her father's, which complicates the analysis.

- More knowledge can be integrated to the final classifier to deal with the more sensitive cases such as half-siblings and identical twins, to improve its robustness.

- The problem of traditional identification can be incorporated as well, to yield a classifier able to perform both identification and kinship assessment. This can be a very sensitive approach, considering the problem of identical twins stated in the previous item.

- A natural step forward in this research will be addressing the multi-class problem derived by others degrees of kinship, *e. g.*, parent-child, parent-grandchild, etc (smaller kinship coefficients, as defined in Chapter 2). For this problem, another interesting point to be considered is how factors as gender and age influence a kinship predictor, and possible approaches to alleviate such influences.

- In addition to kinship classification, a score can be attributed to a pair of individuals, based on the likelihood of them being related. This related to a regression

of the characteristics of the couple from a kinship perspective, which has not been addressed in the literature yet.

- Once a robust and comprehensive classifier is trained to deal with different kin relationships, in addition to classifying whether two people are related or not, it can determine the exact degree of kinship of the pair of individuals. This will consist in labeling the sample as belonging to the class corresponding to similar descriptive vectors, for instance.

Acronyms

ASMs Active Shape Models. 15, 23

CoALBP Co-Occurrence of Adjacent LBPs. 26, 28, 29

CV Cross-Validation. 57, 75

DB database. 16, 19, 70

DBCs Dissimilarity-Based Classifications. 44

DT Delaunay Triangulation. 48

FS Feature Selection. 9, 43, 44, 48, 50, 57, 59, 61, 76, 78, 80, 81, 85

HP Human Panel. 53, 82, 84

IDS Individual Dataset. 13, 20, 32, 50, 56, 61

IS Image Set. 53

KKT Karush-Kuhn-Tucker. 41, 42

KNN k-Nearest Neighbors. 7, 71, 75

KV Kinship Verification. 29, 77, 79

LBP Local Binary Pattern. 26–29, 72

LM Lagrange Multipliers. 39

mRMR min-Redundancy Max-Relevance. 44, 61, 84, 85

OMC Optimal Margin Classifier. 38

OpenCV Open-Source Computer Vision library. 49

PCA Principal Component Analysis. 17, 25, 48, 59

PDF Pair Dataset Frontal. 62, 63

PDS Pair Dataset. 45, 49, 50, 53, 56, 57, 59, 61, 62, 78–81

PSM Pictorial Structure Model. 70

QP quadratic programming. 38

RBF Radial Basis Function. 57, 70

RDF Random Decision Forest. 75, 76

RIC-LBP Rotation Invariant Co-occurrence among Adjacent LBPs. 26, 29, 80

RIGF Rotation Invariant Gabor Feature. 30, 80

SDK Software Development Kit. 55

SFS Sequential Forward Selection. x, 44, 61, 62, 70, 85

SIFT Scale-Invariant Feature Transform. 30, 49, 50, 72, 73, 84

SMO Sequential Minimal Optimization. 41

SoA State of the Art. 3

SURF Speeded Up Robust Features. 49, 50

SVM Support Vector Machines. 7, 33, 41, 43, 45, 47, 57, 62, 63, 68, 70, 71, 75, 76, 78, 82–85

TSL Transfer Subspace Learning. 8

WLD Weber Local Descriptor. 68, 71–75, 78, 80, 84

Bibliography

[1] M. Pantic and L. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," *Proceedings of the IEEE*, vol. 91, pp. 1370–1390, sept. 2003.

[2] Y. Fu, G. Guo, and T. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1955–1976, nov. 2010.

[3] A. Bottino and A. Laurentini, "The analysis of facial beauty: An emerging area of research in pattern analysis," *Lecture Notes in Computer Science*, vol. 6111/2010, pp. 425–435, 2010.

[4] T. F. Vieira, A. G. Bottino, and M. DeSimone, "Automatic kinship recognition." http://areeweb.polito.it/ricerca/cgvg/kin.html.

[5] A. Laurentini, A. Bottino, M. DeSimone, and T. Vieira, "A new problem in face image analysis: Finding kinship clues for siblings pairs," in *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods*, no. 119, pp. 405–410, 2012.

[6] T. F. Vieira, A. Bottino, A. Laurentini, and M. Simone, "Detecting siblings in image pairs," *The Visual Computer*, pp. 1–13, 2013.

[7] T. F. Vieira, A. Bottino, and I. U. Islam, "Automatic Verification of Parent-Child Pairs from Face Images," in *XVIII Iberoamerican Conference of Pattern Recognition (CIARP)*, (Havana, Cuba), 2013.

[8] W. D. Hamilton, "The genetical evolution of social behaviour. i and ii," *J Theor Biol*, vol. 7, pp. 1–16, Jul 1964.

[9] J. N. Bailenson, S. Iyengar, N. Yee, and N. A. Collins, "Facial similarity between voters and candidates causes influence," *Public Opinion Quarterly*, vol. 72, no. 5, pp. 935–961, 2008.

[10] G. Kaminski, S. Dridi, C. Graff, and E. Gentaz, "Human ability to detect kinship in strangers' faces: effects of the degree of relatedness," *Proc Biol Sci*, vol. 276, pp. 3193–3200, Sep 2009.

[11] J. Park, M. Schaller, and M. Van Vugt, "The psychology of human kin recognition: Heuristic cues, erroneous inferences, and their implications," *Review of General Psychology*, vol. 12, no. 3, pp. 215–235, 2008.

[12] L. M. DeBruine, F. G. Smith, B. C. Jones, S. C. Roberts, M. Petrie, and T. D. Spector, "Kin recognition signals in adult faces," *Vision Research*, vol. 49, no. 1, pp. 38–43, 2009.

[13] S. Biswas, K. Bowyer, and P. Flynn, "A study of face recognition of identical twins by humans," in *Information Forensics and Security (WIFS), 2011 IEEE International Workshop on*, pp. 1 –6, 29 2011-dec. 2 2011.

[14] M. F. Dal Martello and L. T. Maloney, "Where are kin recognition signals in the human face?," *Journal of Vision*, vol. 6, no. 12, pp. 1356–66, 2006.

[15] R. Fang, K. D. Tang, N. Snavely, and T. Chen, "Towards computational models of kinship verification," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pp. 1577–1580, Sept. 2010.

[16] G. Somanath and C. Kambhamettu, "Can faces verify blood-relations?," in *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2012.

[17] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? metric learning approaches for face identification," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 498 –505, 29 2009-oct. 2 2009.

[18] S. Xia, M. Shao, and Y. Fu, "Kinship verification through transfer learning," in *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Three*, IJCAI'11, pp. 2539–2544, AAAI Press, 2011.

[19] M. Shao, S. Xia, and Y. Fu, "Genealogical face recognition based on UB KinFace database," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pp. 60–65, June 2011.

[20] G. Guo and X. Wang, "Kinship Measurement on Salient Facial Features," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, pp. 2322 –2325, 2012.

[21] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models – their training and application," *Comput. Vis. Image Underst.*, vol. 61, pp. 38–59, Jan. 1995.

[22] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proceedings of the 10th European Conference on Computer Vision: Part IV*, ECCV 2008, (Berlin, Heidelberg), pp. 504–513, Springer-Verlag, 2008.

[23] S. Milborrow, "Active Shape Models with Stasm." http://www.milbo.users.sonic.net/stasm/.

[24] A. Bottino and S. Cumani, "A fast and robust method for the identification of face landmarks in profile images," *W. Trans. on Comp.*, vol. 7, pp. 1250–1259, Aug. 2008.

[25] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Am. A*, vol. 4, pp. 519–524, Mar 1987.

[26] R. Nosaka, C. H. Suryanto, and K. Fukui, "Rotation invariant co-occurrence among adjacent lbps," in *Proceedings of the ACCV2012 Workshop LBP2012*, pp. 1–11, 2012.

[27] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision amp; Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1, pp. 582–585 vol.1, 1994.

[28] R. Nosaka, Y. Ohkawa, and K. Fukui, "Feature extraction based on co-occurrence of adjacent local binary patterns," in *Advances in Image and Video Technology* (Y.-S. Ho, ed.), vol. 7088 of *Lecture Notes in Computer Science*, pp. 82–91, Springer Berlin Heidelberg, 2012.

[29] F. Bianconi, A. Fernández, and A. Mancini, "Assessment of rotation-invariant texture classification through Gabor filters and Discrete Fourier Transform," in *Proceedings of the 20th International Congress on Graphical Engineering (XX INGEGRAF)*, (Valencia, Spain), June 2008.

[30] K. E. a. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1582–96, Sept. 2010.

[31] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, Nov. 2004.

[32] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT Descriptor with Color Invariant Characteristics," *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, vol. 2, pp. 1978–1983, 2006.

[33] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[34] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[35] A. Ng, "Lecture Notes on Machine Learning." http://cs229.stanford.edu/materials.html.

[36] R. T. Rockarfeller, *Convex Analysis*. Princeton University Press, 1970.

[37] L. Wolf and S. Bileschi, "Combining variable selection with dimensionality reduction," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, pp. 801 – 806 vol. 2, june 2005.

[38] A. Y. Ng, "Feature selection, l1 vs. l2 regularization, and rotational invariance," in *Proceedings of the Twenty-First International Conference on Machine learning*, ICML-04, (New York, NY, USA), pp. 78–, ACM, 2004.

[39] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226 –1238, aug. 2005.

[40] A. W. Whitney, "A Direct Method of Nonparametric Measurement Selection," *IEEE Transactions on Computers*, vol. C-20, no. 9, pp. 1100–1103, 1971.

[41] E. Pekalska and R. P. Duin, *The Dissimilarity Representation For Pattern Recognition: Foundations and Applications*. Singapore: World Scientific Publishing, 2005.

[42] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF : Speeded Up Robust Features," in *Computer Vision-ECCV 2006*, pp. 404–417, 2006.

[43] "Open Source Computer Vision library." http://opencv.org.

[44] C. GmbH, *FaceVACS-SDK Version 8.3 C++ Reference Manual*. Cognitec GmbH.

[45] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe, "Frvt 2006 and ice 2006 large-scale experimental results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 831 –846, may 2010.

[46] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[47] R. Yan, "MatlabArsenal: A MATLAB Package for Classification Algorithms." http://archive.is/kIdhF, 2006.

[48] G. Somanath, M. Rohith, and C. Kambhamettu, "Vadana: A dense dataset for facial image analysis," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pp. 2175–2182, 2011.

[49] J. Lu, "Neighborhood repulsed metric learning for kinship verification," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '12, (Washington, DC, USA), pp. 2594–2601, IEEE Computer Society, 2012.

[50] S. Xia, M. Shao, J. Luo, and Y. Fu, "Understanding Kin Relationships in a Photo," *IEEE Transactions on Multimedia*, vol. 14, pp. 1046 –1056, 2012.

[51] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, "Wld: A robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1705 –1720, sept. 2010.

[52] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, pp. 137–154, 2004. Springer Verlag.